Proceedings of the 53rd Hawaii International Conference on System Sciences | 2020

Preventing Algorithmic Bias in the Development of Algorithmic Decision-Making Systems: A Delphi Study

Banu Aysolmaz
Maastricht University
b.aysolmaz@maastrichtuniversity.nl

Nancy Dau
Maastricht University
n.dau@student.maastrichtuniversity.nl

Deniz Iren
CAROU Open Universiteit Heerlen
deniz.iren@ou.nl

Abstract

In this digital era, we encounter automated decisions made about or on behalf of us by the so called Algorithmic Decision-Making (ADM) systems. While ADM systems can provide promising business opportunities, their implementation poses numerous challenges. Algorithmic bias that can enter these systems may result in systematical discrimination and unfair decisions by favoring certain individuals over others. Several approaches have been proposed to correct erroneous decision-making in the form of algorithmic bias. However, proposed remedies have mostly dealt with identifying algorithmic bias after the unfair decision has been made rather than preventing it. In this study, we use Delphi method to propose an ADM systems development process and identify sources of algorithmic bias at each step of this process together with remedies. Our outputs can pave the way to achieve ethics-by-design for fair and trustworthy ADM systems.

1. Introduction

Nowadays, we increasingly encounter automated decisions that are made regarding or on behalf of us in many aspects of our lives. Businesses implement ADM systems to provide better products and services to their customers through the use of machine learning [1] and deep learning [2], or generally, artificial intelligence (AI) [3]. As a result, ADM systems take over decision-making that affects individuals' daily lives in various ways, including the criminal justice system, education, private and public finance, healthcare, housing, the legal sector, marketing, policies, recruiting, and social media [4, 5, 6, 7, 8]. While ADM systems provide promising business opportunities, their implementation poses several challenges. One of the main concerns of ADM systems is algorithmic bias [9]. In the process of developing ADM systems, bias could be introduced to the system. This can potentially lead to socially and ethically serious consequences [10]. In case of algorithmic bias an ADM system could, for example, fail to act in objective fairness and discriminate systematically and unfairly [11] by favoring certain individuals or groups over others [12]. Ethical concerns for ADM systems are being highlighted by institutions and governments (e.g., [9, 13, 14]). Existence of bias is a significant obstacle in front of the "*principle of justice*" as part of the ethical concerns, which articulates that decisions made by algorithmic systems shall be fair and equal for all human beings [9].

A number of studies have focused on correcting biased decisions made by ADM systems, for example by focusing on algorithmic awareness [15, 16], algorithmic accountability [17, 18, 19], and algorithmic transparency [20, 21]. However, it seems that remedies against algorithmic bias in literature have mainly dealt with detecting bias after decision-making rather than preventing it in the first place. Algorithmic bias revealed after decision-making can bring huge costs to businesses in the form of lost customer image and regulatory charges [22]. Therefore, it is important for businesses to discover and eliminate the sources of bias during the development and deployment of ADM systems. Ethics-by-design, preventing the existence of bias and discrimination through implementing proper technical capabilities in the system development [23], is a potential approach to achieve this goal. Researchers have until now mostly discussed only at a conceptual level about how to realize ethics-by-design [24].

In information systems, various frameworks are used to structure the life cycle of a systems development project. These frameworks provide the best practices about the process to be followed for developing a system. Organizations use these frameworks to ensure that the developed system achieves its expected results in the most efficient way. Although frameworks have been developed for data mining and data science systems, ADM systems have not been specifically handled yet [25, 26]. Defining the process of ADM system development may allow the designers of these systems to better understand how biases are introduced

URI: https://hdl.handle.net/10125/64390
978-0-9981331-3-3

in the development phase. Investigating the whole ADM system development from the earliest design phase is a way to achieve ethics-by-design and is a key to realizing trustworthy ADM systems [9]. Therefore, in this study we aim to define an ADM system development process and identify sources of algorithmic bias at each step of this process. To achieve this goal, we set up a Delphi study with nine experts who have expertise in ADM systems. We prepared an initial questionnaire using four relevant process frameworks, namely CRISP-DM, ASUM, DMLC, and TDSP. In three rounds, the experts developed an eight-step process for ADM system development and identified the sources of bias and related remedies in each of these steps. The outputs were finalized with interviews. The outputs of this study, ADM system development process and the sources and remedies of bias for each step of the process, can be used by businesses to eliminate algorithmic bias while ADM systems are developed and deployed.

2. Literature Review

2.1. Algorithmic bias

One of the ethical concerns about ADM systems is the existence of algorithmic bias among others such as privacy, anonymity, and misuse of data and model [10]. The use of algorithms should lead to fairer decisions since algorithms are objective and not inherently biased [3]. ADM systems should not discriminate people based on sensitive personal data including but not limited to any data that reveals a person's "racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, ..., genetic data, biometric data, ..., data concerning health or sex life or sexual orientation" [27]. Still, algorithmic bias exists. It can be defined as "discrimination that is systemic and unfair in favoring certain individual groups over others" [12]. Algorithmic bias leads to ADM systems to not act in objective fairness but to systematically and unfairly discriminate [11] because of the inherent nature of categorical data. The way ADM system developers collect, select, prepare, and use the data to train the algorithms can introduce bias into the ADM system even when there is no discrimination intention [5, 28, 29].

ADM systems are developed by following a certain process and bias can potentially enter the ADM system during each step of this process. Bias can be introduced due to various factors such as poorly selected and unrepresentative datasets, pre-existing bias inherent in ADM system coders, or technical limitations in the design. The European Commission [9] identifies the collection and selection of training data, which is used to train the ADM system, as one of the ways for bias to enter the system. More generally, humans designing the ADM systems may be a salient reason of bias being introduced to these systems [9, 30]. An example is the COMPAS recidivism prediction software, which is used to predict the probability of a defendant in a criminal case to re-offend and help judges make sentencing decisions [31, 22]. The recidivism scores were found to be biased against African-Americans [32]. In the cases of Google [15] and LinkedIn [33, 34], high-paying job ads were more frequently shown to men than to women.

Incidents of algorithmic bias support the societal and ethical concerns for ADM systems. These concerns become apparent because decisions made through algorithms are consequential for us as they decide, for example, whether we get approved for a loan [31, 35, 22, 36, 16] or get accepted for a job [5, 37, 19]. The societal and ethical problem with algorithmic bias is that our sensitive personal data should not have any effect on how a decision is made and that biases of any kind should not influence decision-making. Nonetheless, because ADM systems are modeled by humans and are based on data provided by humans who are not free from bias and prone to error, ADM systems involuntarily inherit human biases and, thus, human influences are embedded in these systems [38, 30]. Since the goal of using ADM systems is to achieve objective, data-driven, and fair decision-making and decision-making by ADM systems have a growing impact on our lives, algorithmic bias has to be diminished if not entirely eliminated [9].

2.2. Remedies against algorithmic bias

As discussed in the above section, ADM systems can discriminate based on sensitive personal data. Removing such data, however, does not correct unfair ADM systems. To begin with, omitted variable bias tells us that excluding a variable is insufficient to avoid discrimination as any remaining variables that correlate with the excluded variable still contain information about the excluded variable [39]. Similarly, even when sensitive personal information is excluded from the data, algorithms may still discriminate based on this information due to the correlations existing in the overall data [5]. Current research found that to ensure that ADM systems do not discriminate, for instance with respect to race, information about race needs to be used when modeling the algorithms [19]. These finding are corroborated by [40] who not only argue that sensitive personal data are often needed to inspect whether algorithms discriminate but further add that when sensitive information is used responsibly, the discrimination can be made transparent.

Several approaches have been proposed to correct erroneous decision-making due to algorithmic bias. Table 1 provides an overview of research areas that consider managing algorithmic bias in ADM systems. Research in these areas address challenges inherent to ADM systems, such as algorithmic bias, through increasing awareness, accountability, transparency, and auditability of ADM systems. These studies have either an outsiders' perspective to deal with algorithmic challenges, or deal with algorithmic bias "after-the-fact". There is limited literature on overcoming bias or other ethical concerns as part of the ADM development process [10]. Eliminating algorithmic bias has been focus of non-academic entities as well. On the one hand, firms such as Facebook [47], Google [48], IBM [49, 50, 51], and Microsoft [52, 53] have launched tools aimed at detecting and examining AI bias [54]. On the other hand, the European Commission has worked on the "Ethics Guidelines for Trustworthy AI" [9]. According to the EU, one of the key points in designing AI is to incorporate the requirements of trustworthy AI early in the design phase [9]. Thus, there exists a need to systematically look into the ADM system development process with an "insider" perspective to understand possible sources of bias "ex-ante". In the next section, we describe our research to fulfill this need.

Table 1. Review of research areas on managing algorithmic bias

Research Area	Solution to algorithmic bias
Algorithmic awareness	- Enhance public awareness and knowledge about ADM systems to empower users [17, 28] - Increase awareness of users [17] - Raise awareness to possible biases inherent in ADM systems [15, 16]
Algorithmic accountability	- Hold accountable not the system itself but its developers for the decisions [17, 18, 19] - Put in place mechanisms that compensate individuals who have fallen victim to erroneous decision making [17] - Investigate black boxes through algorithmic accountability reporting [41]
Algorithmic transparency	- Make algorithms public [20, 21, 42] - Make ADM systems more understandable and less complex [43, 44] - Reduce algorithmic bias by detecting errors in input data which resulted in unfair decision-making [15]
Algorithm audit	- Develop auditing methods for third parties to determine what algorithms are doing and whether or not algorithm providers are expected to be good or evil [45, 46]

3. Research Method

3.1. Delphi study

Delphi study is a method for structuring a group communication process. It is performed in multiple rounds of questionnaires moderated by a coordinator within the rounds [55]. Since experts do not directly face each other, bias is reduced that can exist when diverse groups of experts meet together. Studies in information systems have used Delphi study as a forecasting tool to make predictions about characteristics of future technologies [57] as well as to help with the introduction, selection, and application of new technologies [58]. Delphi study is most useful for achieving consensus in the areas characterized by uncertainty and lack of empirical evidence [56]. Therefore, using Delphi method in this study is suitable for us to elicit experts' tacit knowledge and generate new ideas on a complex and rapidly-changing domain.

3.2. Procedure for selecting experts

Two important aspects to panel selection are the qualifications of the experts and the panel size [60]. It is essential that experts are chosen based on their expertise and commitment to multiple rounds of questions. We identified potential panelists based on expertise and interest in the fields of ADM systems. Depending on the scope of the problem and available resources, the panel size may vary, which have started from around ten in previous studies [60]. To ensure randomness, we sent invitations to a large group of experts (50) identified via purposeful sampling of professionals that indicated experience about ADM or related concepts on their online profiles. These experts were selected from the organizations within professional network of the authors. Nine experts accepted the invitation. The panelists should constitute a heterogeneous group in terms of background [56] to reduce single-culture bias and provide diverse insights. The panelists in the study included four academics and five practitioners. The panelists of three different nationalities (Dutch, German, and Slovak) worked in seven different industries (information and technology services, education, automobile, telecommunications, logistics and supply chain, auditing) with diverse roles (professor, PhD candidate, data analyst, IT consultant). The experience years on ADM systems was 2.6 on average, varying between 0.5 to 11. When the experts accepted the invitation, we provided them with information about the purpose of the study, how much time it requires, what they need to provide, and what will be done with the information. Afterwards, the experts received the link to the first survey.

3.3. Questionnaire development

The purpose of the first survey was to initiate the process of generating ideas and to give the experts a starting point for their thoughts. The first

survey serves to structure data collection to avoid the randomness and chaos of open-ended dialogs [56]. Using a process framework is imperative to explore how bias or other ethical considerations can emerge during system development [10]. However, there is no accepted framework yet for data science projects [25]. We identified four related process frameworks, as summarized in Table 2. These frameworks represent current standards in the fields related to ADM systems, such as Data Science and Data Mining and, hence, can be used as a starting point to develop an appropriate framework for ADM systems. Hence, we used these frameworks to prepare the first survey as described in the below sections.

Table 2. Chosen frameworks for the first survey in Delphi Study

Framework	Description	Steps
CRISP-DM (Cross-Industry Standard Process for Data Mining)	Process model for carrying out data mining projects [61]	1. Business Understanding 2. Data understanding 3. Data preparation 4. Modeling 5. Evaluation 6. Deployment
ASUM (Analytics Solutions Unified Method)	Extended version of CRISP-DM, which integrates agile principles [62]	Project Management accompanies all steps 1. Analyze 2. Design 3. Configure & Build 4. Deploy 5. Operate & Optimize
DMLC (Data Mining Life Cycle)	Approach to managing and optimizing data mining projects [63]	1. Business Understanding 2. Data understanding 3. Objectives/Hypotheses 4. Select/Sample 5. Pre-Process 6. Transformation 7. Data mining 8. Evaluation 9. Deployment
TDSP (Team Data Science Process)	Methodology that provides a lifecycle for the development of data science projects [64]	1. Business Understanding 2. Data Acquisition & Understanding 3. Modeling 4. Deployment 5. Customer Acceptance

3.4. Data collection and analysis

The recommended number of rounds in a Delphi study is two or three, since more rounds may result in a slower convergence among expert opinions [58]. We designed the study in three rounds each lasting one week. During each round, experts filled out an online survey, as described in the sections below. The study was conducted between December 2018 and January 2019. Experts were given five days to fill in the survey. The following two days were used to qualitatively analyze and aggregate the responses, and design the next survey. The second and third surveys were designed based on prior survey responses and the member checks. Inputs from all experts were taken into consideration and all suggestions were included in the subsequent surveys even if just one expert made a suggestion. In each survey, experts were asked if they agreed with the changes that have been made. In case experts disagreed, they had the opportunity to voice that opinion in the survey. There was only one case, which is further described in the results, where ideas opposed. This issue was resolved during member checks. By showing experts each others' ideas and reasoning, they eventually settled on one idea. All questions were open-ended. This allowed for more qualitative comments to be made and increased the richness of the collected data [60]. Prior to sending out the link of the subsequent survey, we conducted member checks to reduce researcher bias [65]. In these checks, we clarified any ambiguities and verified our understanding of the comments with each expert. The design of each survey is desribed below.

First Survey: The objective was to make the experts develop a new ADM system development process. To do that, experts were introduced the four chosen frameworks and asked to evaluate which steps of these frameworks they deem relevant for ADM systems, and explain their choice. Then, they were required to describe their own process for ADM system development. After analyzing all responses, we aggregated the results and drafted the first version of the ADM system development process.

Second Survey: To expand and finalize the process, the experts were presented with a first draft of the process. They were instructed to indicate whether they find the process complete. If they did not, the experts were asked to explain which steps should be changed. Next to that, the experts were asked to describe how biases could potentially enter ADM systems for each step separately.

Third Survey: To finalize the process, the experts were shown the second draft of the process and were asked whether they agree with the process's steps and feedback loops. If they disagreed, they were asked to explain what should be changed and why. Additionally, the experts were provided with the list of biases that could enter ADM systems during the process, which was prepared based on responses to the second survey. For each step separately, the experts were instructed to indicate whether they agree with the list of biases, and if not, what biases they find redundant or missing from the list. Afterwards, the experts were asked how these biases could be best reduced or eliminated.

Interviews: We interviewed two experts to discuss the final version of the process, the list of potential biases and their remedies.

4. Results

We collected the responses through Qualtrics survey tool. For each round, we downloaded the responses, edited, and analyzed through coding. After analyzing the previous survey's responses, these responses were used as the basis for the subsequent survey.

4.1. Intermediate Results

We first describe how the ADM process evolved in three stages below.

First Survey: To design their own ADM process, the experts referred to the four frameworks. With qualitative content analysis we searched for, sorted, and categorized key steps and related tasks mentioned by the experts Similar responses were combined and overlapping ones were removed. Overall, experts found parts or all of the four frameworks relevant for ADM systems. Three out of nine experts stated that their own process would look similar to the CRISP-DM, two said the same for the TDSP, and one for the DMLC. Moreover, one expert identified the overall project management and another one the involvement of the end user as important for ADM systems. Based on these responses, we drafted the first version of the ADM System Development Process. This version consisted of the following steps taken from the indicated framework:

(1) Project Management (ASUM), (2) Business Understanding (CRISP-DM), (3) Data Acquisition (TDSP), (4) Data Understanding (TDSP), (5) Data Preparation (CRISP-DM), (6) Modeling (CRISP-DM), (7) Evaluation (CRISP-DM), (8) Deployment (CRISP-DM).

Second Survey: While two experts thought that the process was complete, others made suggestions to adjust it. First, data selection was suggested to be a phase together with data acquisition. This emphasizes the importance of selecting the right data to analyze, i.e., thinking about data needs before further data is acquired for the development of the system. Second, a feedback loops between data selection & acquisition, data understanding and data preparation were added to emphasize the iterative structure of these steps. Third, evaluation and testing were divided into two separate steps to accentuate their different purposes and tasks performed accordingly. Fourth, a feedback loop between evaluation and data selection & acquisition was added to highlight the dependency between these two steps. Fifth, a feedback loop between deployment and testing was added to introduce an agile way of working. Sixth, customer involvement was included as a task of project management.

Third Survey: A consensus level of 50% has been aimed for [66] and was achieved during this round. There was an improved degree of agreement compared to the previous round. In the last survey, six experts thought that the process was complete while three indicated that it needed adjustments. During the member checks, we discussed the responses as well as the reasoning behind these suggestions. Based on the responses, following changes were made to the framework. First, a feedback loop between modeling and business understanding was added. This aimed to emphasize the trade-off between time invested in the project and optimality of the ADM system. According to the pareto rule, 80% of the obtainable results should be achieved with 20% of the time. If the customer desires more optimality, additional time and money is needed. Second, the feedback loop from evaluation to data selection & data acquisition was redirected to business understanding. In the previous round, it was questioned whether one would go all the way back to business understanding if something went wrong during evaluation. However, at this round, it was argued that it makes sense to go back to business understanding to make sure that the business objective was correctly understood or to discuss the feasibility of the objective with the customer. This issue was resolved during the member check. As a result, the feedback loop from evaluation to business understanding was re-established. Third, the feedback loop from deployment to testing was deleted because once the ADM system is already deployed, data analysts would not go back to the testing phase to perform simulations of the real world.

Interviews: During the two interviews we presented the experts with the raw list of biases to double check their analysis with ours. In both times, the experts decided to delete most of the same biases as we did. The experts' opinions have been taken into consideration and confirmed with other experts.

4.2. ADM System Development Process

At the end of three rounds, ADM system development process depicted in Figure 1 was developed. It describes the general process of how ADM systems are developed, including steps, related tasks, and loops and reworks among the steps. It also serves as a blueprint for conducting an ADM system project. Since there are various types of ADM systems, the framework is flexible and needs to be customized to these systems' differing purposes. Eight steps have been identified as critical and necessary to build a comprehensive ADM system, each composed of several tasks. The arrows and feedback loops represent

important and frequent relationships between steps. The outer circle symbolizes the cyclic nature of the whole process. The process starts anew once an ADM system is deployed. Lessons learned can lead to a new, and possibly better business understanding so that new projects can benefit from previous ones. While not a step itself, project management is an important component overarching the whole process.

4.3. List of Algorithmic Biases

Using a qualitative approach of coding, we searched the list of potential biases provided by the experts for common thematic elements, grouped them into categories, reassigned biases to more suitable steps, and deleted responses. Then, we assigned the remedies to related biases as described by the experts. These biases and remedies, as shown in Table 3, are organized by steps and common themes which were identified earlier in the coding process.

Biases were found to potentially enter the ADM system during all steps. Table 3 shows that the greatest potential for bias to enter the system is during the phases of business understanding, data selection & acquisition, data understanding, and data preparation. More specifically, we identified the following common themes in biases based on our coding:

- Miscommunication and countercheck,
- Lack of business/technical expertise, miscommunication, and business objective,
- Reason behind gathering data, data collection, data characteristics, data analysis, and time,
- Problem-solving,
- Testing environment,
- Results, and
- Biased project members and maintenance.

5. Discussion and Conclusion

This study aimed to define an ADM system development process and identify the biases that can be introduced to ADM systems during this process together with relevant remedies. Through the outcomes of this study, we aim to enable organizations developing ADM systems to eliminate or decrease algorithmic bias in ADM systems during their development. This constitutes a step through achieving ethics-by-design in ADM systems, which is crucial to achieve fair and trustworthy ADM systems [9].

The resulting ADM system development process not only supports the notion that the way ADM system developers collect, select, prepare, and use the data to train the algorithms can introduce bias into the ADM system [5, 28, 29] but also expands it by introducing other ways of bias entering ADM systems. Our findings suggest that biases can possibly enter ADM systems at any step. However, most critical issues emerge in the initial steps, namely business understanding, data selection & acquisition, data understanding, and data preparation. The further the progress of the project, the less potential there is for bias to enter the system. Since the business understanding step represents the starting point of the whole process, it is critical to identify how biases can be introduced to the ADM system early onwards. Selecting and acquiring data, then understanding that data before preparing it for modeling is an iterative process within the ADM system development process. Biases that can enter the system during these three steps have mutual themes, such as the input of data in these steps, that cut across these steps. Thus, the most common causes of algorithmic bias relate to a faulty business understanding and biased data. The latter supports the assertion that the collection and selection of training data are predominant ways for bias to enter the system [28]. Consequently, algorithmic bias can be substantially reduced during the first four steps. Nonetheless, miscommunication and failure to counter-check by supervisors and other team members facilitate bias to enter during any step since project management oversees all steps. Another finding is that the risk of algorithmic bias is found equally high across all steps but further biases relating to the modeling, testing, evaluation, and deployment steps have not been encountered yet or have not been mentioned in this study. Furthermore, this study corroborates the findings by [19] and [40] since the proposed process does not suggest excluding sensitive personal information.

5.1. Implications

From a research perspective, this work complements existing methods for tackling the issue of algorithmic bias. Previous studies have examined this issue through increasing awareness, accountability, transparency, and auditability of ADM systems. However, challenges with ADM systems, such as the lack of transparency [45], make it difficult for outsiders to investigate ADM systems [21]. Therefore, this paper suggests that ADM system developers and deployers themselves can reduce algorithmic bias, rather than using tools to detect and examine bias [54]. The outcomes of this study can give ADM system developers and deployers the opportunity to prevent algorithmic bias from occurring altogether by eliminating bias by design.

From a practical perspective, eliminating bias is especially important for businesses who may have already experienced incidents of algorithmic bias that

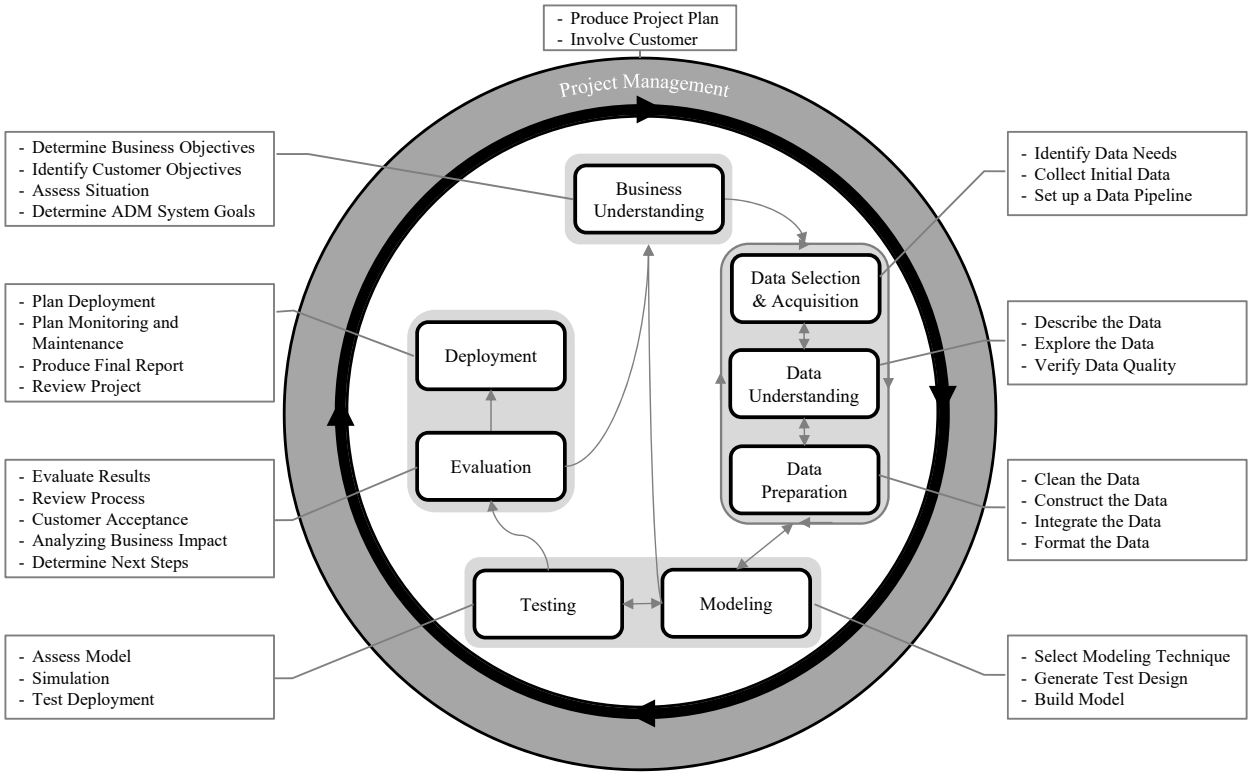

Figure 1. Final ADM system development process

may result in severe or unintentional consequences of decision-making. In such cases, algorithmic bias could lead to societal and ethical problems when, for example, ADM systems discriminate against people of color [31] or against women [32, 15, 11]. Likewise, the presence of algorithmic bias could lead to missing the goal of deploying ADM systems in business operations in the first place. If the ultimate goal of using an ADM system is objective, data-driven, and fair decision-making, algorithmic bias must be prevented. By consulting the process and bias list of this study, IT managers can gain insights into when biases are most likely to enter the system and which remedies could help eliminate them. Businesses should take advantage of the process's flexibility by customizing it to their specific ADM system. Other potential benefits for practitioners include improved quality in project management, a way of benchmarking methodologies and the overall adoption of ADM systems, and increased understanding for ADM systems and their pitfalls.

5.2. Limitations and Future Directions

The first limitation relates to the research methodology. While there are numerous reasons to choose the Delphi study, weaknesses of this methodology include the lack of clarity regarding the means by which consensus may be defined and the resultant differing interpretations [60]. The methodology has also been criticized for forcing consensus and not allowing panelists to discuss issues [67]. Another limitation relates to the research design. Open questions in the surveys allowed us to collect in-depth insights from the experts. This required experts to spend significant time, which may decrease the motivation and quality of the outputs. For this, we used tools to guide experts' thought process, such as presenting related frameworks. Since a common discussion environment among all experts has not been established, such as in a focus group study, the results of the study may be subject to researcher bias. To eliminate this bias, we conducted member checks to verify our understanding with the experts. Additionally, the number of experts may have limited the identified sources of algorithmic bias or remedies. The inclusion of other experts, for example those that work on mission-critical applications, may have resulted in other ideas, such as redundant system development by multiple teams as a potential remedy.

This paper represents a starting point for researchers to look deeper into how ADM system developers and deployers themselves can eliminate algorithmic bias.

Since there are various types of ADM systems, this study can be replicated for specific ADM systems to cater to their distinct characteristics and purposes. The development of an ADM system development process is an important contribution of this study, since lack of a defined process has been found as an important challenge for the success of big data projects [25]. It should be noted that this process was developed with a focus on algorithmic bias. This process can be used as a baseline and extended to consider the introduction of other ethical concerns in the development process, such as privacy, data accuracy, and data misuse [10]. The applicability of the process and listed biases should also be evaluated in a real-life context. Other potential research areas are about the investigation of sources of bias beyond the development of a single system. Even for an ADM system that seems to be robust to bias at deployment, new risks can be introduced during the use and maintenance of that system, e.g., through additional training data. Furthermore, the interaction of multiple ADM systems in a decision chain may uncover unexpected biases that are not apparent for a single system.

6. References

References

[1] P. Adler, C. Falk, S. A. Friedler, T. Nix, G. Rybeck, C. Scheidegger, B. Smith, and S. Venkatasubramanian, "Auditing black-box models for indirect influence," *Knowledge and Information Systems*, vol. 54, no. 1, pp. 95–122, 2018.

[2] B. Goodman and S. Flaxman, "European union regulations on algorithmic decision-making and a "right to explanation"," *AI Magazine*, vol. 38, no. 3, pp. 50–57, 2017.

[3] Royal Academy of Engineering, "Algorithms in decision-making," tech. rep., 2017.

[4] S. Barocas, S. Hood, and M. Ziewitz, "Governing Algorithms: A Provocation Piece." 2013.

[5] F. Bonchi, C. Castillo, and S. Hajian, "Algorithmic bias: from discrimination discovery to fairness-aware data mining," pp. 1–7, 2016.

[6] Federal Trade Commission, "Big Data - A Tool for Inclusion or Exclusion." 2016.

[7] K. Hannah-Moffat, "Algorithmic risk governance: Big data analytics, race and information activism in criminal justice debates," *Theoretical Criminology*, pp. 1–18, 2018.

[8] Science and Technology Committee, "Algorithms in decision-making." 2018.

[9] European Comission, "Draft Ethics Guidelines for Trustworthy AI," Tech. Rep. December, 2018.

[10] J. S. Saltz and N. Dewar, "Data science ethical considerations: a systematic literature review and proposed project framework," *Ethics and Information Technology*, vol. 21, no. 3, pp. 197–208, 2019.

[11] M. Sears, "AI Bias And The 'People Factor' In AI Development," 2018.

[12] A. Springer, J. Garcia-Gathright, and H. Cramer, "Assessing and Addressing Algorithmic Bias — But Before We Get There." 2018.

[13] The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, "Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems," tech. rep., 2019.

[14] Y. Bengio, "The Montréal Declaration: Why we must develop AI responsibly," 2018.

[15] A. Datta, M. C. Tschantz, and A. Datta, "Automated Experiments on Ad Privacy Settings," *Proceedings on Privacy Enhancing Technologies*, vol. 2015, no. 1, pp. 92–112, 2015.

[16] L. Sweeney, "Discrimination in Online Ad Delivery." 2013.

[17] C. of experts on internet intermediaries, "Algorithms and Human Rights Study on the human rights dimensions of automated data processing techniques and possible regulatory implications." 2018.

[18] M. Rouse, "What is algorithmic accountability?," 2017.

[19] I. Žliobaitė and B. Custers, "Using sensitive personal data may be necessary for avoiding discrimination in data-driven decision models," *Artificial Intelligence and Law*, vol. 24, no. 2, pp. 183–201, 2016.

[20] N. Diakopoulos and M. Koliska, "Algorithmic Transparency in the News Media," *Digital Journalism*, vol. 5, no. 7, pp. 809–828, 2017.

[21] R. Kitchin, "Thinking critically about and researching algorithms," 2016.

[22] S. Newell and M. Marabelli, "Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of 'datification'," *Journal of Strategic Information Systems*, vol. 24, no. 1, pp. 3–14, 2015.

[23] V. Dignum, "Ethics in artificial intelligence: introduction to the special issue," *Ethics and Information Technology*, vol. 20, pp. 1–3, mar 2018.

[24] V. Bonnemains, C. Saurel, and C. Tessier, "Embedded ethics: some technical and ethical challenges," *Ethics and Information Technology*, vol. 20, no. 1, pp. 41–58, 2018.

[25] J. Saltz, N. Hotz, D. Wild, and K. Stirling, "Exploring project management methodologies used within data science teams," in *Americas Conference on Information Systems 2018: Digital Disruption, AMCIS 2018*, pp. 1–5, 2018.

[26] J. S. Saltz and I. Shamshurin, "Big data team process methodologies: A literature review and the identification of key factors for a project's success," in *2016 IEEE International Conference on Big Data (Big Data)*, pp. 2872–2879, dec 2016.

[27] Parliament and Council of the European Union, "Regulation (EU) 2016/679 of the European Parliament and of the Council ," *Official Journal of the European Union*, vol. L119, pp. 1–88, 2016.

[28] European Commission, "Algorithmic Awareness-Building," 2018.

[29] E. James, "Practical Steps to Addressing Algorithmic Bias," 2018.

[30] D. Boyd and K. Crawford, "Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon," *Information Communication and Society*, vol. 15, no. 5, pp. 662–679, 2012.

[31] P. Adler, ·. C. Falk, S. A. Friedler, ·. T. Nix, G. Rybeck, C. Scheidegger, ·. B. Smith, and ·. Suresh Venkatasubramanian, "Auditing black-box models for indirect influence," *Knowledge and Information Systems*, vol. 54, pp. 95–122, 2018.

[32] J. Angwin, J. Larson, S. Mattu, and L. Kirchner, "Machine Bias," 2016.

[33] M. Day, "How LinkedIn's search engine may reflect a gender bias," aug 2016.

[34] H. Reese, "Bias in machine learning, and how to stop it," nov 2016.

[35] S. Hajian, *Simultaneous Discrimination Prevention and Privacy Protection in Data Publishing and Mining*. Phd, Universitat Rovira i Virgili, 2013.

[36] T. Speicher, H. Heidari, N. Grgic-Hlaca, K. P. Gummadi, A. Singla, A. Weller, and M. B. Zafar, "A Unified Approach to Quantifying Algorithmic Unfairness: Measuring Individual & Group Unfairness via Inequality Indices," in *Proceedings of the 24th ACM SIGKDD KDD '18*, vol. 10, pp. 2239–2248, ACM, 2018.

[37] F. Kamiran, F. Kamiran, and I. Žliobait, "Quantifying explainable discrimination and removing illegal discrimination in automated decision making," *Knowl Inf Syst*, vol. 35, pp. 613–644, 2013.

[38] K. Kirkpatrick, "Battling algorithmic bias," *Communications of the ACM*, vol. 59, no. 10, pp. 16–17, 2016.

[39] K. A. Clarke, "The Phantom Menace: Omitted Variable Bias in Econometric Research," *Conflict Management and Peace Science*, vol. 22, pp. 341–352, 2005.

[40] B. A. Williams, C. F. Brooks, and Y. Shmargad, "How Algorithms Discriminate Based on Data they Lack: Challenges, Solutions, and Policy," *Source Journal of Information Policy Journal of Information Policy*, vol. 8, no. 8, pp. 78–115, 2018.

[41] N. Diakopoulos, "Algorithmic accountability reporting: On the investigation of black boxes," *A Tow/Knight Brief*, pp. 1–33, 2014.

[42] S. Vijayakumar, "Algorithmic Decision-Making," *Harvard Political Review*, jun 2017.

[43] B. Berendt and S. Preibusch, "Better decision support through exploratory discrimination-aware data mining: foundations and empirical evidence," *Artif Intell Law*, vol. 22, pp. 175–209, 2014.

[44] B. Lepri, N. Oliver, E. Letouzé, A. Pentland, and P. Vinck, "Fair, transparent, and accountable algorithmic decision-making processes," *Philosophy & Technology*, vol. 31, no. 4, pp. 611–627, 2018.

[45] C. Sandvig and K. Hamilton, "An Algorithm Audit." 2014.

[46] C. Sandvig, K. Hamilton, K. Karahalios, and C. Langbort, "Auditing Algorithms : Research Methods for Detecting Discrimination on Internet Platforms," *Communications of the ACM*, vol. 93, no. 1, pp. 1–23, 2014.

[47] D. Gershgorn, "Facebook says it has a tool to detect bias in its artificial intelligence," may 2018.

[48] J. Wexler, "The What-If Tool: Code-Free Probing of Machine Learning Models," 2018.

[49] IBM, "AI and bias - IBM Research," 2018.

[50] Z. Kleinman, "IBM launches tool aimed at detecting AI bias," sep 2018.

[51] K. Varshney, "Introducing AI Fairness 360." 2018.

[52] W. Knight, "Microsoft is creating an oracle for catching biased AI algorithms," *MIT Technology Review*, p. 25, may 2018.

[53] Robitzski, "Microsoft Announces Tool To Catch Biased AI Because We Keep Making Biased AI," 2018.

[54] P. Teich, "Artificial Intelligence Can Reinforce Bias, Cloud Giants Announce Tools For AI Fairness," *Forbes*, p. 24, sep 2018.

[55] V. Mahajan, H. A. Linstone, and M. Turoff, "The Delphi Method: Techniques and Applications," *Journal of Marketing Research*, vol. 13, no. 3, p. 317, 1976.

[56] A. L. Delbecq, A. H. Van de Ven, and D. H. Gustafson, *Group techniques for program planning : a guide to nominal group and Delphi processes*. Glenview: Scott, Foresman and Company, 1975.

[57] R. El-Gazzar, E. Hustad, and D. H. Olsen, "The Journal of Systems and Software Understanding cloud computing adoption issues: A Delphi study approach," *The Journal of Systems and Software*, vol. 118, pp. 64–84, 2016.

[58] D. Gallego and S. Bueno, "Exploring the application of the Delphi method as a forecasting tool in Information Systems and Technologies research," *Technology Analysis and Strategic Management*, vol. 26, no. 9, pp. 987–999, 2014.

[59] M. Patton, "Qualitative evaluation and research methods." 1990.

[60] C. Powell, "The delphi technique: myths and realities," *Journal of Advanced Nursing*, vol. 41, no. 4, pp. 376–382, 2003.

[61] C. Shearer, "The CRISP-DM Model: The New Blueprint for Data Mining," *Journal of Data Warehousing*, vol. 5, no. 4, pp. 13–22, 2000.

[62] IBM, "Analytics Solutions Unified Method - Implementations with Agile principles." 2016.

[63] M. Hofmann and B. Tierney, "Development of an Enhanced Generic Data Mining Life Cycle (DMLC)," *The ITB Journal*, vol. 10, no. 1, pp. 50–71, 2009.

[64] G. Ericson, W. A. Rohm, J. Martens, C. Casey, B. Harvey, S. Gilley, and K. J. Poulton, "What is the Team Data Science Process?," 2017.

[65] K. K. Franklin and J. K. Hart, "Idea generation and exploration: Benefits and limitations of the policy delphi research method," *Innovative Higher Education*, vol. 31, no. 4, pp. 237–246, 2007.

[66] R. Schmidt, K. Lyytinen, M. Keil, and P. Cule, "Identifying Software Project Risks: An International Delphi Study," *Journal of Management Information Systems*, vol. 17, no. 4, pp. 5–36, 2001.

[67] S. Keeney, F. Hasson, and H. P. Mckenna, "A critical review of the Delphi technique as a research methodology for nursing," *International Journal of Nursing Studies*, vol. 83, pp. 195–200, 2001.

Table 3. Sources of algorithmic bias per process step and suggested remedy

Phase	Source of algorithmic bias	Remedy
Project Management	Miscommunication: Poor management	- Use project management software to facilitate communication and planning - Obtain project management certifications
	Counter-check: Lack of double checking by supervisor or other team member	- Let project members double check each other's work - Supervisor should oversee the whole project and has to be able to comprehend all decisions being made
Business understanding	Lack of business/technical expertise: - Misunderstanding the business objective, customers' needs and desires, management directives, corporate implications - Insufficient knowledge about ADM systems	- Increase mutual understanding between business and data people - Make sure the business problem is understood correctly - Make data and business people work closely together to determine the right approach
	Miscommunication: - Incomplete information that results in an incorrect project base - Data analyst's tunnel vision vs stakeholders' diverse points of views	- Ensure written communication - Perform frequent meetings - Make use of disclaimers
	Business objective: - Difficulty of constructing and defining the target variable - Being unaware of the pitfalls of algorithms	- Consider different ways of defining the target variable - Decision trees/distribution analysis/ANN
Data Selection & Acquisition/ Data Understanding/ Data Preparation	Reason behind gathering data: - Aimlessly gathering data - Collecting data based on failures in business understanding	- Take into account the business objective - Go back and forth between phases to make sure the right data will eventually be collected
	Data collection: Technique, origin, purpose	- Implement data collection systems - Digitize the workflow
	Data characteristics: - Unreasonable assumptions regarding data and data quality - Unrepresentative data set - Incorrect granularity	- Verify assumptions with other stakeholders - Gather as much data as possible before choosing from that data set what is most representative and appropriate for the end use - Perform trial and error
	Data analysis: - Unjustified treatment of null variables or outliers - Misinterpreting values in unstandardized data - Differing understandings of unstructured data	- Avoid replacing null variables with the mean or a zero - Avoid deleting null variables and outliers with no reason - Ensure that data analysts are best equipped to correctly understand and interpret data - Implement a fail-safe
	Time: - Data that is not representative for the present/future - Data spans over a long period of time	- Only use the data that is most relevant for the present/future
Modeling	Problem-solving: - Heurisics - Trade-offs between optimality, completeness, accuracy, precision, and robustness - Over-fitting model	- Ensure that the heuristics applied is known and understood by all stakeholders - Perform 80/20 rule - Use train test-splits or cross-validation
Testing	Testing environment: - Failure to test under different scenarios or horizons with different/more current data - Forgetting that results represent predictions and not the reality	- Use a testing catalogue for a structured description of instances and functions to test the system - Perform a sensitivity analysis - Try to predict future data and use it to do simulations
Evaluation	Results: Failure to look out for results that are socially unacceptable or undesirable for the business	- Communicate socially unacceptable or undesirable results to all stakeholders - Go back to business understanding and make sure the that everyone has understood the business objective - Consult experts outside of the project team
Deployment	Biased project members: Having only project members deploy the system	- Consult certified experts outside of the project to deploy the systems in order to avoid the amplification of biases that have not been eliminated in earlier phases
	Maintenance: Failure to retrain the model periodically and to check whether the results can be replicated in the present	- Ensure that the system is calibrated and re-estimated regularly - Build in control mechanisms - Always see the system in the context of the present time