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

Technical Report for the Intermediate Results of the Study

Maastricht University

Appendix A: Copy of Invitation

Invitation to Participate in a **DELPHI STUDY**

Best Practices in **INTERNAL AUDITING OF THE ADM SYSTEMS**

Thank you for you taking your time to review this information. You represent a substantial part in the business culture. You are invited to participate in this study because of your significant expertise and experience in the field of ADM systems.

Algorithmic Bias in Algorithmic Decision-Making (ADM) Systems

With the increasing use of artificial intelligence and all its subcategories in the business context, algorithms are increasingly being used to **automate the process of decision-making**. While this opens up many opportunities, the problem with algorithmic decision-making (ADM) systems is that they can exhibit algorithmic bias. On the one hand, ADM systems can help us overcome our limited information processing capacity and the bounded rationality of humans. On the other hand, if **algorithmic bias** is inherent in ADM systems, this can lead to discrimination that is unfair in favoring certain individuals or groups over others. **There are many possibilities in which biases can enter the ADM system** when we look at the process by which ADM systems are developed. This is mainly because the decision-making of these systems reflect implicit values of the humans who were involved in developing these systems. In other words, 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. Since ADM systems lack transparency because businesses treat algorithms as proprietary trade secrets, said businesses are in the best position to reduce algorithmic bias. One way to do this is through the **internal auditing of the process of how ADM systems are developed**.

The Delphi Study

The purpose of this study is to gather best practices in internal auditing of ADM systems. Specifically, the goal is to develop the *"ADM System Developing Process*" framework which businesses can use to perform internal audits of their ADM systems and further to reduce algorithmic bias. The study will consist of **three rounds, each lasting one week**. In each round, you will fill out an online survey anonymously. This should **take about 10-15 minutes**. After each round, I will consolidate all answers and will send you a new survey. More detailed instructions will be given before each survey.

The following represents an overview of the study.

1) The first round will consists of a semi-structured brainstorming session. You will be presented with a starting point upon which you can expand the framework or develop a different one.
2) The second round reviews a summary of the prior brainstorming session. The consolidated answers of all experts are sent out to you in another survey for you to narrow down the issues you deem most important.
3) After adopting the framework and the first review, a final feedback session will help fine-tune the development of the framework.

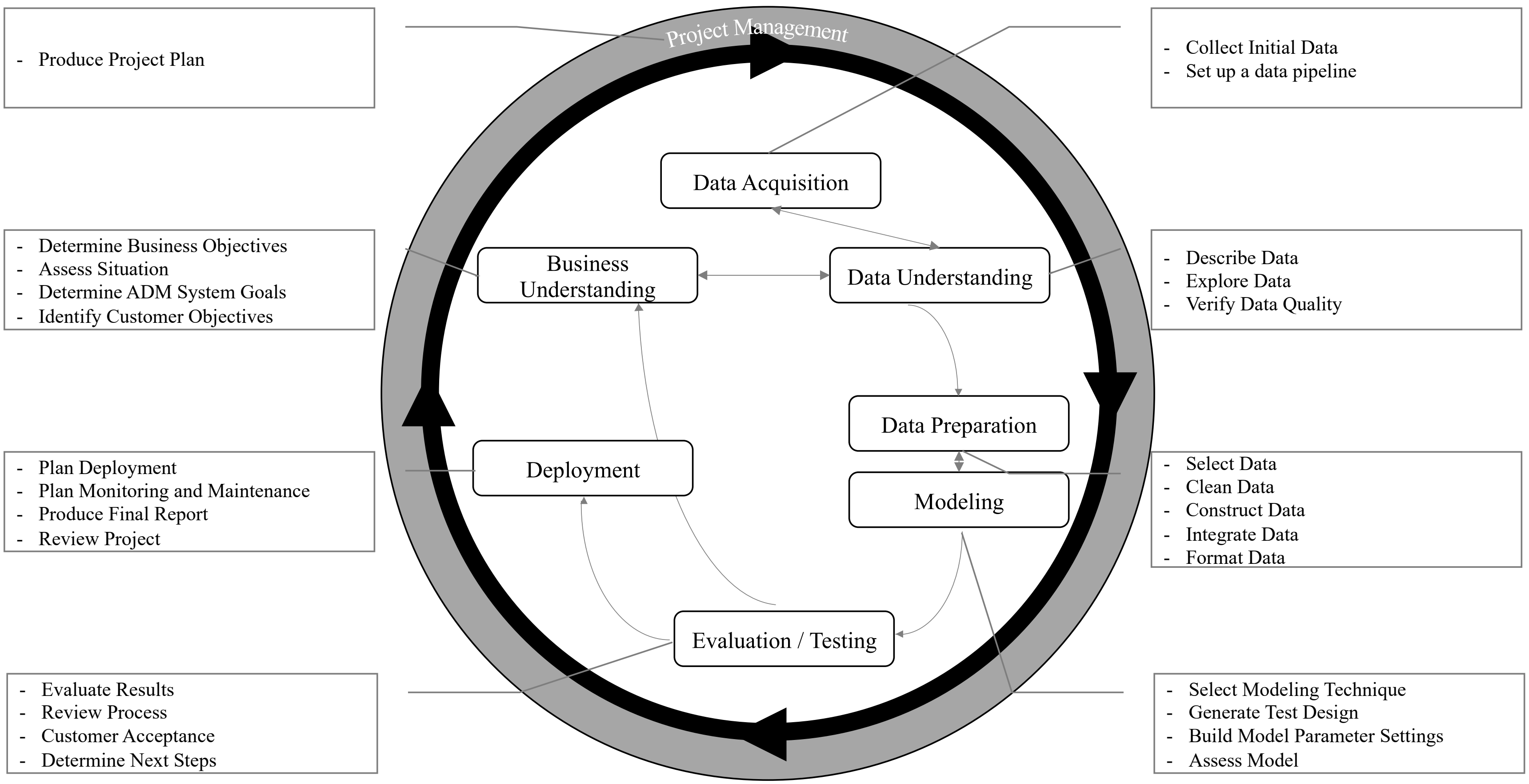
Benefits

Naturally, the final results will be shared with you. By participating in the study you will gain access to the best practices of other businesses and academia. The study aggregates knowledge of all experts and, thus, can facilitate benchmarking of methodologies and the overall adoption of ADM systems.

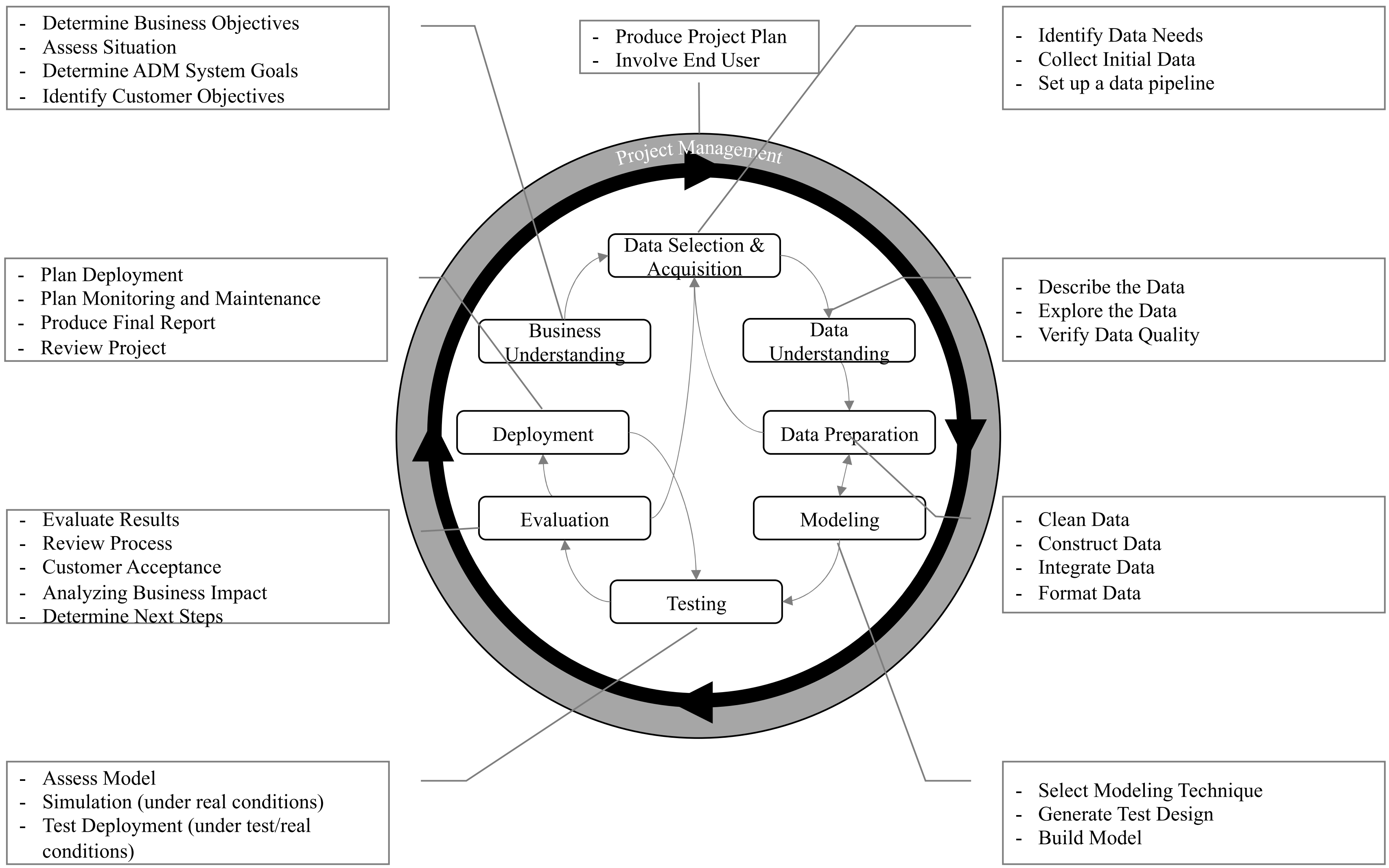
Contact Details

- Nancy Dau
- Tongersestraat 53
- 6211 LM Maastricht
- Maastricht University
- Emails: dau.nancy@googlemail.com, n.dau@student.maastrichtuniversity.nl

Appendix B: First Draft of the ADM System Developing Process Framework



Appendix C: Second Draft of the ADM System Developing Process Framework



Appendix D: List of Biases (unanalyzed)

Phase	Bias
Project Management	1) Time or resource constraints may steer the entire study in a different direction 2) Poor management result in misunderstandings by project members 3) Different point of views from developers and clients 4) Communication problems 5) Pushing for results and not accepting the possibility of finding no significant relationship in the data 6) Not thinking about algorithmic bias
Business Understanding	1) Misinterpreting customers' needs and desires because by default, analysts approach every problem from the same angle, unless that angle does not work 2) Most data scientists do not have a business background and, therefore, do not completely comprehend management directives and corporate implications 3) Constructing and defining the target variable is a difficult task 4) Lack of information etc. can result in an incorrect base of the final project/product 5) Human misunderstanding, communication problems between parties 6) Wrong business objectives 7) Not knowing what the pitfalls of the use of algorithms are 8) No alignment between goal of ADM system use and tactical/strategic plan 9) Only looking at money and time 10) Not identifying kinds of bias that can arise
Data Acquisition	1) Data too sparse or too dense with information 2) Gathering data without knowing what information is precisely needed will results in the model adjusting to fit the data rather than the data being adjusted according to the model's needs 3) How the data is gathered can result into bias, e.g. mode of gathering, origin, reason, purpose, reliability 4) System failure 5) Datasets may be biased 6) Acquiring wrong or incomplete data based on a biased business understanding 7) Collecting false data due to failures in business understanding 8) Human failure in collecting datasets, e.g. data is biased or not enough data is collected 9) Only looking at the data that is collected in the past and not thinking about the possibility of starting to collect data that is actually needed in the present 10) Not having the right granularity 11) Not analyzing trend in data over time 12) Assuming that the data is collected in the same way over time 13) Collecting data that is not representative of the end use

Phase	Bias
Data Understanding	1) Lack of standardized data can result in values being interpreted differently 2) Lack of a proper fail safe to make sure that even when you do not understand the data completely, you make reasonable assumptions so that your model will do what it needs to do 3) Data is mostly unstructured rather than structured so that data can be misinterpreted 4) Wrong connection between variables 5) Wrong interpretation of data 6) Not knowing the actual meaning of the variables and its values 7) Not knowing the true meaning of an empty field or a zero 8) Wrong assumptions 9) Not knowing the demographics of the sample 10) Assuming that the data is not manipulated 11) Exploration bias
Data Preparation	1) Incorrect/harsh assumptions about the data 2) Data is still corrupted after cleaning it and no one notices it 3) Neglecting to manipulate the data in appropriate ways 4) Data is not free of errors and contains null values 5) Replacing empty values with the mean 6) Removing records with empty values 7) Unjustified removal of outliers 8) Preparing the data so that it fits the model 9) Looking at the whole dataset over time and rather than at the most recent dataset that may be more representative 10) Selecting bad data
Modeling	1) Heuristic bias can occur when the problem is too large to be solved to optimality by general methods 2) Lack of supervising managers that oversee the modeling phase 3) Lack of understanding in statistics 4) Data are not adequately prepared and might be biased so that these biases are reproduced by the modeling 5) Class imbalance problem 6) Human failure in building/choosing the model can lead to development failures 7) Only looking at the accuracy of the model and not at other factors like robustness 8) Choosing bad models
Testing	1) Not testing the results as broad as possible by changing or simulating the test instance
Evaluation	1) Misinterpretation of results 2) Ignoring results that are not socially accepted 3) Ignoring results which are undesirable for the company 4) Ignoring the actual business objective

Phase	Bias
Deployment	1) People who were involved in any of the other previous steps should not be the same people who deploy the ADM system to ensure that any possible biases that occurred during any of the previous steps aren't enhanced 2) Predictive model does not cater to the real world 3) Not including maintenance of the model by retraining the model periodically to check whether the results in the past still hold today