The Devil is in the Data.

Datafficational Problems and Solutions applied to the Dutch Childcare Benefits Affair.

Bachelor Thesis Philosophy of a Specific Discipline.
Author: Emma Ligthart
Number of words: 10.501
Supervisor: Prof. Dr. J. de Mul
Advisor: Prof. Dr. F.A. Muller
Main study: Econometrics and Operations Research
Date of completion: 14-06-2022

Table of Contents

Glossary

Toeslagenaffaire	=	Childcare benefits affair, Dutch scandal in which the tax authorities (*Belastingdienst*) wrongly accused childcare benefit receivers of fraud.[1]
Belastingdienst	=	Dutch tax authorities.
Autoriteit Persoonsgegevens	=	Dutch data protection authority. The independent administrative body that is responsible for the supervision of personal data processing in (amongst others) the Dutch tax authorities.

[1] Although the literal translation of *Toeslagenaffaire* is Childcare benefits affair, the word affair is a significant euphemism for the scandal it comprises. However, due to the common use of the Dutch word *Toeslagenaffaire*, this thesis follows the literal translation "Childcare benefits affair."

Introduction

The role of data in our lives has never been greater. From the minute we wake up to the moment we go to sleep data-driven applications and tools help us execute our daily tasks. Think of communication tools such as e-mail or WhatsApp, transportation apps such as Uber, or financial technologies that allow us to pay even with our (smart) watch. In addition, (big) data have had numerous successful implementations in fields such as cancer diagnosis, traffic optimisation and entertainment. Data thus plays a major role in our lives.

However, there is darker side to the ongoing tendency to capture things in data and the consequential importance of data insights. This darker side becomes apparent in data-driven decision making. Apart from helping us execute daily tasks, data also control us. Data have become the most important input in decision making. As adequately captured by Horkheimer and Adorno, it appears as if ever since modernity

"*Anything which cannot be resolved into numbers, and ultimately into one, is illusion.*"[2]

This darker side of data can be advanced more clearly through the concept of datafication. In 2013 Mayer-Schönberg and Cukier popularised the term datafication and gave it the following meaning: the ongoing tendency to (1) capture many aspects of human life in data on a large scale and (2) gain insights from this data.[3] Since its popularisation, there has been an ongoing philosophical debate about the precise content of the term datafication and its consequences.[4] Datafication, the tendency to take insights from data as leading, has lead to numerous catastrophes. The Dutch *Toeslagenaffaire* (Childcare benefits affair) is an important example of such a catastrophe because of its high impact on many families. The use of the problematic risk-classification-model in this affair clearly exhibits the detrimental effects of datafication.

The Childcare benefits affair took place between 2004 and 2019. As part of social security in the Netherlands, parents may request childcare benefit from the government. In attempt to stop fraud regarding this benefit, the Tax authorities implemented the risk-classification-model. This self-learning model established risk profiles of childcare

---

[2] Max Horkheimer and Theodor Adorno, *Dialectic of Enlightenment*, trans. Edmund Jephcott, ed. Gunzelin Schmidt Noerr (Stanford: Stanford University Press, 2002), 4.

[3] Viktor Mayer-Schönberger and Kenneth Cukier, "Datafication," in *Big Data: A Revolution That Will Transform How We Live, Work and Think*, (London: John Murray, 2013).

[4] José van Dijck, "Datafication, Dataism and Dataveillance: Big Data Between Scientific Paradigm and Ideology," *Surveillance & Society* 12, no. 2 (May 2014): 197–208, https://doi.org/10.24908/ss.v12i2.4776; Ulises Mejias and Nick Couldry, "Datafication," *Internet Policy Review* 8, no. 4 (2019): 2, https://doi.org/10.14763/2019.4.1428.

benefit receivers by predicting the probability of fraud. One of the problematic aspects of the model is that it considered foreign nationality as indicative of error or fraud. This resulted in a discriminatory selection of fraudsters. Many innocent families were sued for fraud and wrongfully suffered enormous personal and financial consequences. The Dutch data protection authority (in Dutch *Autoriteit Persoonsgegevens*) fined the tax authorities 2.75 million euros due to their discriminatory and illegal working method.[5]

This example goes to show that datafication can lead to ethically unfair and illegal consequences. For this thesis, I perform a case study of the Childcare benefits affair to constitute the problems and unethical consequences related to datafication in practice. The purpose of the thesis is to present practices that can be implemented to avoid datafication-related (dataficational) problems. I first explain datafication and its four main problems. Next, I present the details of the Childcare benefits affair, focussing on the risk-classification-model. Then I show how the problems of the Childcare benefits affair relate to the four main dataficational problems. Finally, I present solutions that aim to avoid similar catastrophes.

Two main aspects distinguish this thesis from other works on the Childcare benefits affair. Firstly, most reports on the Childcare benefits affair focus on reconstruction: they have a retrospective focus. The main focus of this thesis is to find solutions to avoid similar incidents. Therefore, the thesis has a forward-looking focus. The reconstruction of the Childcare benefits affair in this thesis only serves to constitute solutions to prevent similar catastrophes. Secondly, many reports on the Childcare benefits affair stem from a governmental or data-based point of departure. However, this thesis stems from philosophical premises and takes the concept of datafication as leading. This thesis focusses on the ontological analysis of the concepts of data and datafication and applies the philosophical problems regarding datafication to the Childcare benefits affair. As such, the proposed solutions are related to a broader dataficational tendency within our society.

The thesis is structured as follows. Section 1 explains what datafication is and the four main problems it poses. Section 2 reconstructs the Childcare benefits affair, focussing on the risk-classification-model. Section 3 describes how the four main dataficational problems occur in the Childcare benefits affair. Section 4 provides possible solutions, focussing on prevention of the problems named in Section 3.

---

[5] Autoriteit Persoonsgegevens, "Boete Belastingdienst voor zwarte lijst FSV," published April 12, 2022, accessed May 2, 2022 https://autoriteitpersoonsgegevens.nl/nl/nieuws/boete-belastingdienst-voor-zwarte-lijstfsv#:~:text=Bovendien%20kreeg%20de%20Belastingdienst%20in,keren%20in%20de%20fout%20gegaan.

1. Datafication and its Problems

1.1 Datafication

The concept of datafication first gained prominence when it was mentioned by Mayer-Schönberger and Cukier in their book *Big Data*.[6] The concept of datafication as popularised by Mayer-Schönberger and Cukier means: extracting information from material and quantifying this information into data, with the goal of gaining insights from that data.[7] In many cases it concerns information that no one considered valuable at first sight, but which becomes valuable through scale and quantification.[8]

After its initial introduction, the term datafication underwent changes and at present, researchers are mostly interested in the datafication of *aspects of human life* into data.[9] Mejias and Couldry (2019) even go as far as stating that datafication refers to "The wider transformation of human life so that its elements can be a continual source of data."[10] As such, datafication entails the process which reduces many aspects of *human life* to data, to be employed for manual and automatic, small and large scale analysis. Whenever aspects of human life are reduced to data, datafication differs from the one we find in the natural sciences. Datafication combines two phenomena: on the one hand the transformation of aspects of human life into data, and on the other hand the generation of value from this data.[11] Data is thus produced for usability purposes.[12]

1.2 Data

Understanding the common sense definition of data is necessary to further analyse datafication. Datafication transforms human information into data, but what is "data"? The word data stems from the latin word *dare*, which means "to give." Data is the plural past participle of *dare* and means "givens."

This etymology translates to the present common-sense meaning of data. According to the *OECD Glossary of Statistical Terms*, data are "characteristics or information, usually numerical, that are collected through observation. […] Data is the

---

[6] Mayer-Schönberger and Cukier, "Datafication," 118.

[7] Ibid., 120.

[8] Ibid., 119-120.

[9] Mejias and Couldry, "Datafication," 2.

[10] Ibid., 2.

[11] Ibid., 3.

[12] Ibid.

physical representation of information in a manner suitable for communication, interpretation, or processing by human beings or by automatic means."[13] In the first sentence of the quote, data is equated with information. In the second sentence, data is defined as a *representation* of information. Representation is the expression of a thing through something else, thus data is understood as the expression of information through numerical or textual symbols.

This equation of data with information renders the the common sense understanding of data (the OECD definition) incorrect and problematic. The concept of information comprises much more than data alone.[14] We can distinguish three dimensions of information, of which data constitutes only one: a syntactic dimension, a semantic dimension and a pragmatic dimension.[15] The syntactic dimension of information concerns its form: whether the information is textual, visual or audible. The semantic dimension concerns the meaning of the information. The pragmatic dimension of information concerns the practical implications for the receiver.[16] Data comprises the syntactic dimension of data: the form of information, which in the case of data is numerical. The meaning and implications of data are not automatically implied by the data itself. In fact, data (syntactic dimension) have to be subjected to careful interpretation and intervention to be translated into meaning (semantic dimension) and action (pragmatic dimension).[17] There are problems inherent to this translation into meaning via data processing which are discussed in Section 1.3.2.

To illustrate the tripartite concept of information, take the example of an article in the English news paper that states that it will be hot tomorrow with a lot of sun.[18] The letters on the paper constitute the syntactic dimension of information: the data. The interpretation of this data by a reader constitutes the semantic dimension of information. Only if an english-reading observer interprets the letters on the paper as meaning that it will be hot tomorrow, the semantic dimension of data succeeds. The pragmatic dimension of information comes in when a reader decides to put on sun screen the next day. There thus are conditions to the letters becoming information in its full tripartite sense. For example, if the message is in Chinese and the reader cannot interpret the

---

[13] Organisation for Economic Co-operation and Development, *OECD glossary of statistical terms* (Paris: OECD publishing, 2008), 119.

[14] Jos de Mul, personal communication, May 16, 2022.

[15] Jos de Mul, "The Informatization of the worldview," in: *Information, Communication & Society.* 2, no. 1 (1999), 81, https://doi.org/10.1080/136911899359763.

[16] De Mul, "The Informatization of the worldview," 81.

[17] Boyd and Crawford, "Critical questions for Big Data," 668.

[18] Example inspired by: Jos De Mul, "The Living Sign. Reading Noble from a Biosemiotic Perspective," *Biosemiotics* 14 (May, 2021), 10, https://doi-org.eur.idm.oclc.org/10.1007/s12304-021-09426-y.

message or if the reader does not have sunscreen, the information is not transferred in its full sense.

In short, the common-sense understanding of data is: the numerical expression of information provided to us by observation. The purpose of data is to process it into useful insights for human beings. However, in contrast to the OECD definition, data constitutes only one of the three dimensions of data. Equalling data to information, thereby disregarding the two other other dimensions of information foreshadows the first problem of datafication in Section 1.3.1 which discusses the theory-ladenness of observation. The disregard for the interpretation and pragmatisation of data leads to an objective image of information following from data.

1.3 Problems of Datafication

I base my investigation into the problems of datafication on the characterisation of the ideology behind datafication as formulated by Van Dijck.[19] Van Dijck explains why datafication constitutes a generally accepted method for understanding social behaviour.[20] Data are considered symptoms of people's actual behaviour. This gives rise to a belief in the possibility to predict and monitor people's behaviour through data.[21] According to her, datafication is rooted in an ideology.[22] This ideology guards trust in (a) the objectivity of quantified data, (b) the potential of tracking human behaviour and (c) the agents that utilise and store the data.[23] I establish four main problems regarding datafication.

1.3.1 Problem 1: Data is Viewed as Objective

The ideology behind datafication constitutes a trust in the objectivity of data.[24] The objective-subjective dichotomy I refer to in this paper is the difference between

[19] Van Dijck, "Datafication, Dataism and Dataveillance," 197-207.

[20] Ibid., 198.

[21] Ibid.

[22] Ibid.

[23] Ibid.

[24] Van Dijck, "Datafication, Dataism and Dataveillance," 200; Danah Boyd and Kate Crawford, "Critical questions for Big Data," *Information, Communication & Society* 5, no. 15 (2012): 666-668, https://doi.org/10.1080/1369118X.2012.678878.

impartial (objective) and not impartial (subjective).[25] I employ the definition as given by Arnold Baise, who states that: "Subjective knowledge (or, more correctly, subjective *belief*) is generally described as being partial or biased or prejudiced, or based on arbitrary assumptions or personal feelings; objective knowledge is then inferred by contrast, as impartial or unbiased or not prejudiced, and so on."[26] I argue that data is rather intersubjective: established through common agreement of a group of two or more subjects.

The misleading view of data being objective is expressed by the common-sense understanding of data, exemplified in the OECD definition in Section 1.2. If you understand data as pure, numerical information received from observation, you understand data to be impartial and unbiased. Data being objective is the logical conclusion that follows almost automatically. However, the concept of data in present day society does not meet its common-sense understanding as objective information. The common-sense understanding of data as objective (exemplified by its OECD definition) is incorrect and problematic for two reasons. Firstly, data are not impartial: they are not given, but are constructed. Secondly, data are not "just there," but are produced.

Firstly, data are not given, but selected and constructed for a specific purpose in mind. According to Gitelman, data have to be imagined as data in order to come into existence.[27] This imagination is intersubjective, in the sense that it is agreed upon by several individuals.[28] The construction of data is not objective, because it is not impartial: it is determined by a specific purpose, which determines which data are gathered and in what manner.[29] This purpose is intersubjective because it depends on the needs and wants of specific individuals or organisations. The intersubjectively chosen purpose remains the main driver behind data, determining whether it is relevant and must be used. The purpose of data collection matters because it determines the sphere of justification: different purposes call for different levels of standards.[30] The higher the impact of the purpose of the data collection, the higher the ethical and legal standard should be.

Data are not "just there," but are produced, which involves human choice. The

---

[25] Arnold Baise, "The objective–subjective dichotomy and its use in describing probability," *Interdisciplinary Science Reviews* 45, no. 2 (2020), 181, https://doi-org.eur.idm.oclc.org/10.1080/03080188.2019.1705559.

[26] Baise, "The objective–subjective dichotomy," 181.

[27] Lisa Gitelman, *'Raw Data' is an Oxymoron* (Cambridge: MIT Press, 2011), 3.

[28] Ibid.

[29] Boyd and Crawford, "Critical questions for Big Data," 667.

[30] Jos de Mul, personal communication, May 16, 2022.

production of data entails quantification and abstraction and therefore the omission of certain aspects. Which aspects are omitted is intersubjectively chosen.[31] Also in the process cleaning data, intersubjective decisions are made about the relevance of certain attributes.[32] This makes data inherently intersubjective.

To illustrate this construction and production of data, take the example of a railway company wanting to optimise the efficiency of train travelling, ensuring fast transportation and smooth transfers. It might appear that the data needed for this purpose—number of transfers and travellers per station per day—follows automatically and is therefore impartial, thus objective. However the reality of train travelling is much more complex and other data are relevant as well. What data is considered relevant depends on intersubjective choices. Football fanatics might consider data on the locations and dates of important football matches important. Weather forecasters most likely regard weather data important because more people take the train with bad weather. Climate activists probably find the entire purpose of train efficiency inferior to the purpose of increasing sustainability of train travels. They might consider the data irrelevant, simply because they do not assign importance to its purpose.

With this example, the concept of theory-ladenness of observation enters the conversation on the objective image of data. Theory-ladenness of observation entails that observations, and thus data, are never pure or objective, but are always determined by the theoretical frame through which an observer observes. Rather than the *ding an sich*, we only grasp the *ding für mich*. Observation is determined by theoretical presupposition. Data are never objective because presuppositions determine what is observed: what observations do become and which do not become data. Rather, data are intersubjective: their existence and meaning are determined by the agreement of a group of individuals. Presuppositions of such a group determine whether football match dates or weather forecasts are constructed and produced into data.

In short, data is intersubjective because they are determined by a specific purpose that is agreed upon by a group of individuals, data are selected and constructed, and data involve omission of certain aspects. Data are not simply there, but are gathered for specific purposes and with specific interests. Data are not natural phenomena, but are intersubjectively produced and contain human choice. It is important that the wrongful view of data as objective is abandoned, both for those who provide their data, to determine whether or not they want to distribute their data, as well as for those that work with data, so they can properly determine the value of their work.

---

31 Gitelman, *'Raw Data' is an Oxymoron*, 4; Boyd and Crawford, "Critical questions for Big Data," 667.

32 Boyd and Crawford, "Critical questions for Big Data," 666-668.

1.3.2 Problem 2: Bias, Unexplainability and Intransparency in Data Processing

Trust in the objectivity of data establishes trust in the insights of data. As described in Section 1.3.1, the common sense understanding of data equates data to information. This disregards the necessity of interpretation of data to generate meaning and insight. The common-sense definition of data conveys that gathering data and creating insight from data are one and the same process. However, in reality the construction and the interpretation of data are two separate processes. Data are not raw facts, but are only building blocks that can be put together to form an argument.[33] Data are given to provide a rhetorical basis.[34] In itself data do not provide insights. Interpretation is necessary for insight to arise.

This disregard for the interpretation of data encourages an objective view of *insights* from data and neglects the problems inherent to data processing. The interpretation of data involves human choice and is inherently problematic. The problems inherent to data processing are the central theme of the second problem of datafication. This problem is tripartite: it includes the bias, unexplainability and intransparency of algorithmic data processing.

During the processing of data into insight, human biases are automated and amplified.[35] The *automatisation* of human bias is exemplified by a self-learning recruitment model. During training, the model is provided with historical data: CV's that have been manually labelled as either good or bad candidates. As such, the training data solely entails the biased human judgment of candidates. The model learns from this biased data and therefore automates the human bias inherent to the labelling. Disastrous consequences of automated bias are apparent from the recruitment model implemented by Amazon, which highly favoured men over women.[36]

An example of the *amplification* and *self-reinforcement* of human bias is found in the PREDPOL model.[37] The Dutch police implemented this self-learning model to predict criminal behaviour by registering the place, time and perpetrator of crimes.[38] Based on this data, surveillance was increased in certain neighbourhoods, automatically

---

[33] Gitelman, *'Raw Data' is an Oxymoron*, 7.

[34] Ibid.

[35] Catelijne Muller, "The Impact of Artificial Intelligence on Human Rights, Democracy and the Rule of Law: draft d.d. March 20, 2020," *ALLAI*, March 20 2020, https://allai.nl/wp-content/uploads/2020/06/The-Impact-of-AI-on-Human-Rights-Democracy-and-the-Rule-of-Law-draft.pdf.

[36] Muller, "The Impact of AI on Human Rights," 10.

[37] Jos de Mul, "Metaphors we nudge by," in *Nudging media*, ed. James Katz (London: Palgrame MacMillan, unpublished manuscript, 10 May 2022), 9.

[38] Ibid.

causing more crime arrests within that area. Data of these new arrests were again fed into the model.[39] This new arrest data reestablished the area as high risk because of increased arrests.[40] In short, once a neighbourhood became high risk, more surveillance led to more arrests and this data reinforced the model to mark the neighbourhood as high risk.[41]

Apart from biases present in data processing, algorithmic data processing is often unexplainable. We can divide algorithms broadly into two categories: rule-guided and self-learning.[42] In case of rule-guided algorithms, rules are explicated by humans and the workings of the algorithm are explainable.[43] However in case of self-learning algorithms the workings are unexplainable, even to the data scientists that design the networks.[44] Although unexplainably, self learning algorithms often perform better at tasks than humans, for example in distinguishing benign form malignant tumors or linking genetic abnormalities to diseases.[45] This creates an explainability-performance tradeoff: increased performance is accompanied by unexplainability.

In addition to being biased and unexplainable, algorithmic processing of data is intransparent in two ways. Firstly, the unexplainability of self-learning algorithms creates intransparency about the reasoning behind an algorithmic decision. Algorithms come to be treated as black boxes, causing outcomes of an algorithmic decision making process to be treated as given. This black box effect impedes any challenging or contending of outcomes of an algorithm.[46] Self-learning algorithms are intransparent in a second sense because their use often remains undisclosed by organisations.[47] This impedes control of data usage and thereby touches upon the third problem of datafication.

In short, insights from data are not immanent but are produced through algorithmic data processing. These algorithms automate and amplify human bias and are often unexplainable and intransparent. It is important that the bias, unexplainability and intransparency of data processing is brought to the attention of the general public. In addition, effort in the form of research, regulation or governance should be dedicated to solve this tripartite problem, or at least diminish its impact.

---

[39] Ibid.

[40] Ibid.

[41] Ibid.

[42] Ibid., 7.

[43] Ibid.

[44] De Mul, "Metaphors we nudge by," 7.

[45] Ibid.

[46] Warren von Esbach, "Transparency and the Black Box Problem: Why We Do Not Trust AI," *Philosophy & Technology* 34 (September 2021): 1612, https://doi.org/10.1007/s13347-021-00477-0.

[47] Muller, "The Impact of AI on Human Rights," 15-16.

1.3.3 Problem 3: Difficulties in Control of Data Usage

The third problem of datafication concerns difficulties regarding (governmental) control of data usage. As touched upon in Section 1.3.2., the intransparency and unexplainability of algorithms impede governmental control. Without knowing that certain data or algorithms are used, control is impossible. Transparency about the use of data and algorithms is therefore crucial. In addition, control over the explainability of algorithms is not straightforward because there exists a tradeoff between explainability versus performance. The question becomes how much performance should be sacrificed for explainability. As such, the solution is not as simple as requiring full explainability.

Control is hampered even further by the double role of the government regarding data control. There is a public expectation of straightforward governmental control of data usage by companies. However, in reality the government wears a so-called double hat. On the one hand the government aims to control data usage, but on the other hand it employs data for its own purposes, such as the protection of citizens (for example protection from terrorist attacks).[48] This double hat complicates governmental control of data usage.[49] The government both has to control data usage and shares the interest in data usage with companies and academia to analyse and predict behaviour.[50] The question is how well the controlling force of the government remains when it shares its interests with the thing to be controlled.[51]

The national control of data usage is complicated even further because many data-centred organisations (i.e. Facebook or Google) transcend national boundaries. At present, there is not one institution in command of data control, but the interplay between different institutions predominates.[52] Although legislation from the European Union is approaching with the EU Artificial Intelligence Act, enforcement of these laws at national level is difficult. The fast paced innovation many data-centred companies maintain adds to that difficulty.[53]

---

[48] Jos de Mul, "Hoe de overheid vertrouwen verspeelt - en kan heroveren," *Trouw*, Letter & Geest, 4 juli 2015, 4, https://www.trouw.nl/nieuws/hoe-de-overheid-vertrouwen-verspeelt-en-kan-heroveren~bda40169/.

[49] Ibid.

[50] Van Dijck, "Datafication, Dataism and Dataveillance," 202.

[51] Ibid., 202-204.

[52] Marion Brivot and Yves Gendron, "Beyond panopticism: On the ramifications of surveillance in a contemporary professional setting," *Accounting, Organizations and Society* 36 (2011): 136, https://doi.org/10.1016/j.aos.2011.03.003.

[53] De Mul, "Hoe de overheid vertrouwen verspeelt," 4.

In short, control over data usage is complicated by unexplainability and intransparency, the double hat of the government, and the crossing of national boundaries. The public expects governmental data control, often unaware of its difficulties. Where traditional public phenomena are more easily regulated by national governments, datafication regularly escapes this control. As such, the public might feel more secure than it actually is and consequentially hand out data more liberately than appropriate. There is a need for centralised control regarding datafication.

1.3.4 Problem 4: Dataveillance

The fourth problem of datafication regards the use of predictive analytics by both governments and companies in order to control people. Increasingly popular predictive analytics form future hypotheses based on historical data, thereby enabling dataveillance.[54] Dataveillance is the phenomenon of continuously tracking individuals to predict future incorrect behaviour on the basis of present data.[55] Dataveillance is problematic in two respects.

Firstly, dataveillance no longer judges individuals on the basis of past behaviour, but on the basis of predicted future behaviour. Judgment on the basis of future behaviour is problematic because it violates the principle of innocence and limits narrative causality.[56] The principle of innocence entails that individuals must be presumed innocent until proven guilty. The principle of innocence is violated by presuming guilt based on predicted future behaviour, because guilt is assumed before the action is even executed. Thus, guilt is assumed before it is definitively proven.

Narrative causality is the capacity to motivate actions through reasoning.[57] Predicting future behaviour solely based on past behaviour eliminates narrative causality because no room is left for explaining future behaviour through reasonable motivation.[58] This raises the question if we can be held responsible for predicted future actions if we have no reasonable motivation or motive for them.[59] Apart from the question of responsibility, the loss of our narrative causality impedes our individuality.

---

[54] John Wang, Qiyang Chen and James Yao, "Data Mining Fundamental Concepts and Critical Issues," in *Encyclopedia of Artificial Intelligence*, eds. Juan Dopico, Julian Dorado and Alejandro Pazos, (Hershey: IGI Global, 2008), 418-423.

[55] Van Dijck, "Datafication, Dataism and Dataveillance," 205.

[56] Jos de Mul, personal communication, May 16, 2022.

[57] Jos de Mul, "Help ik ben een database!," *De Groene Amsterdammer*, March 13, 2022, 49, https://www.groene.nl/artikel/help-ik-ben-een-database.

[58] Ibid.

[59] Ibid.

Predictive actions extrapolate past behaviour and therefore don't leave room for individual change or coincidence.[60]

Dataveillance is problematic in a second sense because it effectuates unconscious and thus involuntary tracking and control. Instead of tracking certain individuals for specific purposes, dataveillance tracks all individuals without predetermined purposes.[61] This creates a Foucaultian panopticon, in which not only the guards, but even the guard tower has become invisible.[62] There is no clarity regarding when and how our actions and characteristics are captured in data nor what is done with this data. Although people might agree to the usage of their data for certain purposes, such as the prevention terrorist attacks, it is not precluded that their data is used for other purposes as well. Such other purposes often come down to behavioural control.[63] As such, predictive programming and dataveillance become connected with exploitation: people are unconsciously and involuntarily controlled through the data they themselves produce.[64]

---

[60] Ibid.

[61] Van Dijck, "Datafication, Dataism and Dataveillance," 204-207.

[62] De Mul, "Hoe de overheid vertrouwen verspeelt," 6.

[63] Mark Andrejevic, "Exploitation in the data-mine," in *Internet and Surveillance: The Challenges of Web 2.0 and Social Media*, eds. Christian Fuchs, Kees Boersma, Anders Albrechtslund, and Marisol Sandoval, 71-88, (New York: Routledge, 2012), 72-73.

[64] Ibid.

2. The Childcare Benefits Affair

This section is dedicated to the factual reconstruction of the Childcare benefits affair. An account of the datafícational problems of the Childcare benefits affair is reserved for Section 3.

2.1 Broad Context

The Childcare benefits affair (in Dutch *Toeslagenaffaire*) refers to a collection of harmful policy decisions and actions executed by the Dutch tax authorities (in Dutch *Belastingdienst*) and the deficient attempts to restore the harm by the Dutch government.[65] The defects regard the provision of childcare benefits. Childcare benefits are part of the social security system in the Netherlands. The benefit is provided to assist parents in paying for formal childcare. The tax authorities are responsible for the distribution of childcare benefits.[66]

The benefits-system is highly based on trust, which makes the system vulnerable for fraud. The tax authorities therefore execute their tasks in a field of tension between fraud prevention and service.[67] The prevention of fraud in the social security system has climbed its way up the political agenda since the 1980's.[68] Due to political choices, from 2011 onwards the focus of the tax authorities regarding childcare benefits has shifted to fraud prevention.[69] The anti-fraud policy left little to no room to adjust measures to individual circumstances and caused harsh consequences for many innocent parents.[70]

In 2019, it came to light that the tax authorities wrongly identified almost thirty thousand innocent parents as fraudsters.[71] These parents were unjustly obliged to pay

---

[65] Amnesty International, "Xenofobe machines: Discriminatie door ongereguleerd gebruik van algoritmen in het Nederlandse toeslagenschandaal, " London: Amnesty International Ltd, October 2021, 9, https://www.amnesty.org/en/documents/eur35/4686/2021/nl/.

[66] Autoriteit persoonsgegevens, "Belastingdienst/Toeslagen De verwerking van de nationaliteit van aanvragers van kinderopvangtoeslag," Rijksoverheid, July 7, 2020, 4, https://www.autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/onderzoek_belastingdienst_kinderopvangtoeslag.pdf.

[67] Ibid., 4.

[68] Amnesty International, "Xenofobe Machines," 11.

[69] Autoriteit Persoonsgegevens, "Verwerking van de nationaliteit," 5 ; Amnesty International, "Xenofobe machines," 12.

[70] Parlementaire ondervragingscommissie Kinderopvangtoeslag, "Tweede Kamer der Staten-Generaal, Verslag - Parlementaire Ondervragingscommissie Kinderopvangtoeslag: Ongekend Onrecht," Tweede Kamer, December 17, 2020, 7-14, https://www.tweedekamer.nl/sites/default/files/atoms/files/20201217_eindverslag_parlementaire_ondervragingscommissie_kinderopvangtoeslag.pdf.

[71] Gijs Herderscheê, "Ruim 1.100 kinderen van gedupeerden toeslagenaffaire werden uit huis geplaatst," de *Volkskrant*, October 19, 2021, *https://www.volkskrant.nl/nieuws-achtergrond/ruim-1-100-kinderen-van-gedupeerden-toeslagenaffaire-werden-uit-huis-geplaatst~baefb6ff/.*

back tens of thousands of euro's of received benefits.[72] For many of these parents, this led to poverty and out-of-home placement of their children.[73] This scandal has since been referred to as the Childcare benefits affair.

2.2 Allocation of Childcare Benefits

Whether someone is entitled to childcare benefits, depends on both income and nationality. Applicants with suitable income and Dutch nationality are entitled to childcare benefits.[74] There is a special regulation for applicants that do not possess Dutch nationality and are not to be treated as Dutch citizen based on legal provision. In this case, the benefit is subject to the lawful residence of the applicant. Whether such an applicant is entitled to benefits is dependent on the ground of lawful residence.[75]

Partly for that reason, the tax authorities gathered and stored the (double) nationality of applicants of childcare benefits.[76] The tax authorities retrieved data about nationality from the Dutch population administration, in which basic details about every citizen are stored.[77] The nationality of applicants was stored and used in the umbrella system for childcare benefits, called the benefits provision system.[78] The benefits provision system is used for the allocation and payment of benefits and fraud detection.[79] It consists of several subsystems, one of which is the risk-classification-model. The risk-classification-model has been assessed as discriminatory and constitutes one of the most problematic parts of the tax authorities' policy.[80] Therefore, this thesis focuses on the risk-classification-model.

2.3 Design of the Risk-classification-model

The risk-classification-model is a self-learning algorithmic decision-making model. It determines which applications or adjustments of childcare benefits are most likely to be

[72] Autoriteit Persoonsgegevens, "Verwerking van de nationaliteit," 6.

[73] Amnesty International, "Xenofobe Machines," 14.

[74] Autoriteit Persoonsgegevens, "Verwerking van de nationaliteit," 8.

[75] Ibid.

[76] Ministerie van Financiën, "Gebruik van dubbele nationaliteit," December 18, 2020, 1, https://services.belastingdienst.nl/toeslagen-herstel/gebruik-van-dubbele-nationaliteit/.

[77] Autoriteit Persoonsgegevens, "Verwerking van de nationaliteit," 8.

[78] Ministerie van Financiën, "Gebruik van dubbele nationaliteit," 1; Autoriteit Persoonsgegevens, "Verwerking van de nationaliteit," 10.

[79] Autoriteit Persoonsgegevens, "Verwerking van de nationaliteit," 12.

[80] Autoriteit Persoonsgegevens, "Verwerking van de nationaliteit," 3.

incorrect or fraudulent.[81] It is a self-learning model and is trained with historical examples of manually checked applications and adjustments, labelled as correct or incorrect.[82]

The model considers several variables, called indicators, to determine the risk of an application or adjustment, such as the distance between the childcare location and the place of residence.[83] The indicators are initially determined by data specialists of the tax authorities and subsequently tested on statistical correlation with incorrect applications.[84] The indicators with the highest correlation are inputted into the risk-classification model.[85] Through training, the risk-classification-model determines the importance of each variable by assigning penalty points to every indicator.[86] The model assigns higher penalty points to indicators it considers more relevant for the selection of incorrect or fraudulent applications.

After training based on historical applications, the model is ran every month to determine the risk of new applications. In case an application satisfies an indicator, it is assigned the penalty points belonging to that indicator. The model calculates the total penalty score belonging to an application and scales this value to provide the risk score.[87] The model thus assigns a risk score between zero and one (where one is maximum risk) to each application and adjustment, based on similarity of the variables of a new or adjusted application with the indicators of incorrect applications.[88]

After assigning a risk score to all new applications, the model orders them and provides the applications with the highest risk score to the manual control of employees.[89] Employees are not informed on which basis the model has selected the

---

81 Amnesty International, "Xenofobe Machines," 12; Autoriteit Persoonsgegevens, "Verwerking van de nationaliteit," 14.

82 Alexandra van Huffelen, "Kamerbrief inzake de openbaar making van het risicoclassificatiemodel toeslagen," Rijksoverheid, November 26, 2021, 4, https://open.overheid.nl/repository/ronl-aab59576-16e3-49ce-84f2-d57afc760bbb/1/pdf/kamerbrief-inzake-de-openbaar-making-van-het-risicoclassificatiemodel-toeslagen.pdf.

83 Ibid., 12.

84 Ibid.

85 Ibid., 5.

86 Van Huffelen, "Kamerbrief inzake de openbaar making van het risicoclassificatiemodel toeslagen," 5; Amnesty International, "Xenofobe Machines," 17.

87 Amnesty International, "Xenofobe Machines," 17.

88 Autoriteit Persoonsgegevens, "Verwerking van de nationaliteit," 15; Amnesty International, "Xenofobe Machines," 17.

89 Auditdienst Rijk, "Onderzoeksrapport Toeslaggerelateerde CAF-zaken," Rijksoverheid, March 12, 2020, 12, https://open.overheid.nl/repository/ronl-1c359fdc-4398-44e2-9969-1f353c0c481d/1/pdf/rapport-caf-adr.pdf.

application.[90] Several sources indicate that after the manual checking, these applications and their ratings were inputted into the model again as new training cases.[91]

Between 2013 to 2018, the nationality of applicants was used as one of the indicators for the risk-classification model. More specifically, this indicator distinguished between applicants with Dutch and non-dutch nationality. Non-Dutch nationality of the applicant increased the risk score the model attached to an application.[92] The risk-classification-model was thus biased against persons with non-dutch nationality. In a letter by the Secretary of State for Finance to the Dutch parliament, the Secretary admits that the model may have been biased.[93] The tax authorities presupposed a connection between foreign nationality and fraud, using foreign nationality as indicator for fraud.[94]

The Dutch data protection authority assessed the use of the indicator Dutch/non-dutch nationality in the risk-classification-model as discriminatory. The Dutch data protection authority finds the Dutch/non-dutch nationality indicator wrongful because it is irrelevant for the determination of the legitimacy of the application.[95] The use of nationality would only be justified if it was employed to determine lawful residence.[96] However, the indicator Dutch versus non-Dutch nationality cannot be used to determine lawful residence. An applicant with European nationality may not possess Dutch nationality but may have lawful residence and therefore be entitled to benefit.

The Autoriteit Persoonsgegevens finds that the groups on either side of the indicator do not differ in relevant aspects.[97] Thus, there is no objective justification for the differentiation between applicants with Dutch or non-dutch nationality. Therefore it is unjust and illegitimate to penalise one of the two groups. The Dutch data protection authority concludes that the use of the nationality indicator and therefore the risk-classification-model as a whole was discriminatory between 2013-2018.[98]

---

90 Autoriteit Persoonsgegevens, "Verwerking van de nationaliteit," 14.

91 Rijksoverheid, "Werkbeschrijving Risicoselectie," March 2014, 6-9, https://www.rijksoverheid.nl/documenten/rapporten/2014/03/31/werkbeschrijving-risicoselectie---maart-2014; Rijksoverheid, "Bijlage 16-26," November 26, 2021, 59-61, https://www.rijksoverheid.nl/documenten/publicaties/2021/11/26/overige-bijlagen-bij-kamerbrief-over-openbaar-making-risicoclassificatiemodel-toeslagen.

92 Van Huffelen, "Kamerbrief inzake de openbaar making van het risicoclassificatiemodel toeslagen," 6; Autoriteit Persoonsgegevens, "Verwerking van de nationaliteit," 16.

93 Van Huffelen, "Kamerbrief inzake de openbaar making van het risicoclassificatiemodel toeslagen," 7.

94 Amnesty International, ""Xenofobe Machines,"" 17.

95 Autoriteit Persoonsgegevens, "Verwerking van de nationaliteit," 43.

96 Ibid.

97 Autoriteit Persoonsgegevens, "Verwerking van de nationaliteit," 43-44.

98 Ibid., 49-50.

2.4 Context and Use of the Risk-classification-model

This section describes the context and use of the risk-classification-model. The goal of the usage of the model was to prevent (1) errors in applications and adjustments and (2) fraud.[99]

Every month, the model provided a list of high-risk applications to be checked manually by employees.[100] Once selected for manual control, the applicant underwent a suspension of payment of all childcare benefits.[101] During manual control, employees could mark applicants as fraudsters. On many occasions, this mark was given based on small mistakes, such as a missing signature.[102]

The label of fraudster had serious consequences including the obligation to repay any benefits previously received. Thus, selection for manual control by the risk-classification-model could lead to unjust and disproportionate repayment claims. Being selected by the model was detrimental for both innocent and guilty applicants. The risk of these serious consequences was higher for persons with non-dutch nationality, because they were selected for manual control more often, since they were assigned higher risk scores compared to individuals with the Dutch nationality (as described in Section 2.3).

Many innocent applicants were labelled as fraudsters because employees were encouraged to act based on even the slightest suspicion. The tax authorities were forced to prove the success of the risk-classification model by reclaiming money from suspected fraudsters.[103] The repayments from fraudsters were supposed to cover the expenses of the development and training of the model.[104] In a memo about the risk-classification-model, the importance of the proceeds of the model are even emphasised.[105] Thus, there was an incentive to mark as many applicants as fraudsters as

[99] Amnesty International, "Xenofobe Machines," 16; Parlementaire ondervragingscommissie Kinderopvangtoeslag, “Ongekend Onrecht,” 14.
The tax authorities did not directly describe fraud detection as a goal of the risk-classification model. However, the tax authorities named surveillance as one of the goals of the model and defined surveillance as the identification of fraud cases. Therefore, we can nevertheless consider fraud detection as one of its goals.

[100] Autoriteit Persoonsgegevens, "Verwerking van de nationaliteit," 14.

[101] Amnesty International, "Xenofobe Machines," 12.

[102] Ibid., 13.

[103] Amnesty International, "Xenofobe Machines," 13.

[104] Ibid.

[105] Rijksoverheid, “Bijlage 69-75,” November 26, 2021, 78, https://www.rijksoverheid.nl/documenten/publicaties/2021/11/26/overige-bijlagen-bij-kamerbrief-over-openbaar-making-risicoclassificatiemodel-toeslagen

possible without much regard for the justice of the accusations.[106] When attempting to prove their innocence, applicants were unable to receive sufficient information from the tax authorities. As such, they were unable to provide the correct documentation.[107] This further increased the number of applicants wrongfully accused of fraud.

[106] Ibid., 12-13.

[107] Amnesty International, "Xenofobe Machines," 12-13.

3. The Dataficational Problems of the Childcare Benefits Affair

This Section provides an account of the four dataficational problems occurring in the Childcare benefits affair.

3.1 Problem 1: Data is Viewed as Objective

The first problem of datafication is the misleading presupposition that data is objective (given by pure observation), whereas in reality data is selected and constructed for a human-defined purpose. Data is determined by an intersubjective human-defined purpose, which depends on human preference and prioritisation. Due to the theory-leadenness of observation, our presuppositions determine what observations are turned into data. Data thus comes to be intersubjective itself, as explained in Section 1.3.1. In the Childcare benefits affair, we find that the selection and construction of data for the purpose of error and fraud detection are determined by presuppositions and are therefore not objective. The following two instances show the selection and construction of data as determined by intersubjective presuppositions.

Firstly, the indicators which determine the input data of the model originated from the insight of data scientists. Therefore the determination of these indicators contained the presuppositions of the data scientists, inherent to human insight. The mathematical correlation with incorrect applications determined whether an indicator was definitively used. However, the first source of the indicators—the notion of which indicators could possibly be relevant—was human insight. Thus, the input data of the model was constructed by human choice and contained presuppositions inherent to human choice.

The dark side of these presuppositions becomes apparent in the discriminatory indicator of nationality. The initial discriminatory idea of using nationality as an indicator in the model stems from the human choice of data scientists, not from objective mathematical calculation. This choice presupposes that applicants with non-dutch nationality are more likely to commit fraud or make mistakes and is therefore biased.

Whereas other domains may have installed preventive measures against such discriminatory bias, the presupposed objectivity of data hides the necessity of preventive measures. Although most people—and most data scientists—would be concerned if *employees* were instructed to use foreign nationality as indicator of fraud, these concerns vanished when not employees but an *algorithmic model* was instructed to view foreign nationality as indicative of fraud. This hiatus in judgment of employees versus judgment of algorithmic models can be explained by the prejudice that data is objective. Data inputted into an algorithmic model is viewed as objective and therefore consideration of human error or bias is regarded as irrelevant.

The second instance of selection and construction of data is found in the training data of the risk-classification-model. The training data were labels of manually checked applications. These labels were determined by human choice and therefore contained presuppositions inherent to human choice. The training data of the model were biased by these presuppositions. This touches upon the second problem of datafication, which concerns data processing. The consequences of the bias in the training data are therefore further established in Section 3.2.1.

3.2 Problem 2: Bias, Unexplainability and Intransparency in Data Processing

The second problem of datafication concerns the bias, unexplainability and intransparency of data processing, especially regarding self-learning algorithms. The risk-classification-model was a self-learning algorithm and therefore relates to this second problem of datafication.

3.2.1 Bias

The risk-classification-model automated and amplified human bias. As established in Section 3.1 the training examples and the indicators of the model were subject to human presupposition and therefore contained human bias. As explained in Section 1.3.2, this human bias is amplified and automated by self-learning models. Specifically for the risk-classification-model, the biased consideration of nationality as indicative of error or fraud was automated when nationality became an indicator in the risk-classification-model. The training data of the model was biased, in the sense that human presuppositions in the manual determination of applications as incorrect or fraudulent was reflected in the data. During its training the model also automated this bias.

The model even became self-reinforcing. As stated in Section 2.3, several sources indicate that applications that were outputted as high risk were inputted back into the model as training data after they were manually checked.[108] Because only the applications evaluated by the model as high risk were evaluated, fraud was found only in these applications. As such, the model was affirmed in its evaluation of these applications as risky. As a consequence, the model must have assigned increasingly higher risk to the variables of these applications because these applications were more

[108] Rijksoverheid, "Werkbeschrijving Risicoselectie," 6-9; Rijksoverheid, "Bijlage 16-26," 59-61.

likely to be incorrect or fraudulent. Thus, bias present in manual control of applications was not only automated but also amplified by the risk-classification model.

3.2.2 Unexplainability and Intransparency

Unexplainability and intransparency become apparent when we consider the use and implementation of the model. Unfortunately, the unexplainability of the risk-classification-model is not entirely clear. The degree of unexplainability of a model is determined by the type of artificial intelligence employed (for example neural networks, deep neural networks or support vector machines). Unfortunately the type of artificial intelligence employed in the risk-classification-model remains undisclosed.[109] Therefore, all I can state is that because the model was self-learning, it was unexplainable at least to some extent.

The degree of inexplainability becomes less relevant once we consider the intransparency in the use of the risk-classification-model. Even if the model was explainable, this explanation would be meaningless because the model became completely intransparent with its implementation. The employees that checked the high-risk applications determined by the model were not informed about the workings of the model. Hence, the model turned into an absolute black box. None of the employees were informed about why the model provided them with the applications they had to check. As a consequence, employees took the outcomes of the model as "given." It became impossible for employees to contend model outcomes.

Another consequence of the unawareness of the working of the model is prejudiced judgment. It is probable that employees were more likely to label an applicant as fraudulent because they might have presupposed that there must be a reason why the model outputted this application. If employees were aware of the workings of the model, this would prevent such prejudice and lead to more informed, better, judgement.

3.3 Problem 3: Difficulties in Control of Data Usage

The third problem of datafication concerns difficulties regarding control of data usage. Difficulties in control of unethical data and algorithm usage reveal themselves in the Childcare benefits affair. Because the use of the risk-classification-model remained undisclosed for a long time, it was impossible to regulate it. This intransparency impeded control and lead to very late constitution of the problems of the Childcare benefits affair. The model had been in use for over 5 years before its problems were

[109] To the best of my knowledge.

established. If there would have been transparency about the use of the model, its problems would likely have been established much earlier. This could have saved a lot of damage.

The government plays a double role regarding control of data usage. On the one hand it benefits from data usage, on the other hand it is responsible for its control. This problematic double hat reveals itself in practice with the Childcare benefits affair. The tax authorities responsible for the affair are part of the government. The tax authorities therefore carry responsibility for using data and algorithms in an ethical manner. However, on the other hand the tax authorities benefited from uncontrolled data and algorithm usage. The more data available to the risk-classification-model, the more accurate it became, the more incorrect applications were selected and the more benefits flowed back to the tax authorities. As such, there was an incentive for the tax authorities to break the law regarding data processing. Stricter international regulation could alleviate this problem. I return to the topic of international regulation in Section 4.4.

3.4 Problem 4: Dataveillance

The fourth problem of datafication is dataveillance. Dataveillance is the continuous tracking of individuals to predict faulty behaviour. It constitutes a twofold problem: (1) the violation of the principle of innocence and narrative causality and (2) involuntary tracking and control. Both problems occur in the Childcare benefits affair.

With the risk-classification-model, guilt depended on predicted future behaviour. Individuals were not judged based on past behaviour, but based on characteristics that led to a prediction of future behaviour. Applicants were labelled guilty based on this prediction of future behaviour. They were judged guilty prematurely, which violates the principle of innocence. This judgment based on characteristics left no room for narrative causality. The probability of actions was substantiated by a person's characteristics and not by a person's reasonable motivation.

Furthermore, involuntary tracking and control occur in the Childcare benefits affair. Applicants were not informed about what data was used in the determination of error or fraud. Applicants provided data about their nationality for the purpose of public administration. They were never informed about this data being used for determining error or fraud in childcare benefit applications. Applicants may have unconsciously agreed by agreeing to incomprehensible terms and conditions. However, they never consciously agreed to the use of their data for this purpose. They could not have known that providing their nationality for public administration would lead to discrimination regarding childcare benefit. As such, applicant were unconsciously and involuntarily controlled through the data they themselves produced.

4. Proposed Solutions

This chapter proposes solutions to prevent dataficational problems as revealed in practice in the Childcare benefits affair. This chapter does not intend to provide an exhaustive overview of all possible governance and policy related solutions. It attempts to draw up partially new solutions specific to dataficational problems.

4.1 Increase Awareness through Training

Training should be provided to everyone to (1) repair the image of data as selected and constructed and (2) raise awareness of the bias, unexplainability and intransparency of (self-learning) algorithmic data processing. As established in Section 3.1 and 3.2, the presupposed objectivity of data and the problems of algorithmic data processing are root causes of the Childcare benefits affair. Awareness of these problems would induce a critical attitude amongst the general public and thereby create a safety net against dataficational problems. Such awareness can be achieved through training.

Such training should be provided to everyone because awareness should be raised universally. Contrary to the literary suggestion to solely train data scientists, I argue that training should be provided to both users and providers of data.[110] Awareness of the problems amongst users of data stimulates self-criticism on the data insights and models they themselves create. Data or its insights will no longer be regarded as unproblematic. Recognition of the problems amongst providers of data effectuates empowerment because it creates awareness of the consequences of data provision. If one is informed about the consequences of data provision, one can chose to provide data in a more empowered manner. Increased awareness amongst the general public, data scientists and other employees constitutes the most wide-ranging safeguard to unethical data use.

Not only should training be provided universally, it should start from a young age as well. The wrongful view of data as objective is solved most straightforwardly if this wrongful image does not arise at all. For that purpose, the non-objective core of data should be taught from a young age. Awareness of the problems of data processing from an early age is crucial in the fast-paced data driven world children grow up in. The fact

---

[110] Elizabeth Pike, "Defending Data: Toward Ethical Protections and Comprehensive Data Governance," *Emory Law Journal* 69 (April, 2019): 741, https://scholarlycommons.law.emory.edu/elj/vol69/iss4/2; Ray Eitel-Porter, "Beyond the promise: implementing ethical AI," *AI and Ethics* 1 (October 2021): 76, https://doi.org/10.1007/s43681-020-00011-6; Keng Siau and Weiyu Wang, "Artificial intelligence (AI) ethics: ethics of AI and ethical AI," *Journal of Database Management* 31, no. 2 (April, 2020): 82-83, https://doi.org/10.4018/JDM.2020040105.
Users of data include data scientists and other employees.

that the importance of data will only increase over the upcoming decades amplifies the importance of training early on. Classes on the non-objectivity, or intersubjectivity, of data and the bias, unexplainability and intransparency of data processing should be included in high school programs already. They could for example be included in the philosophy or social studies ("maatschappijleer") program. To properly educate this complex subject, training should continue in higher education and professional life.

4.2 Transparency and Explainability

An important condition to the safety net created by awareness is transparency about the use of data and algorithms. Often, it is impossible to be critical of the use of data and algorithms, because their use is undisclosed.[111] Therefore, transparency constitutes ground zero for ethical use of data and algorithms.

Advantageous transparency is more complex than disclosing as much information as possible. Even if there is transparency about the use of data and algorithms, it often remains incomprehensible. Self-proclaimed transparent organisations often publish all possible information regarding data and algorithm usage.[112] This results in intricate, opaque and unintelligible disclosures.[113] Pike finds good middle ground regarding transparency about data and algorithms. According to her, organisations should only publish understandable information regarding the logic in their decision making process and the future pathways of data (insights).[114] Because it is not always clear where data (insights) end up, uncertainty can be part of this disclosure.[115] She proposes that data practitioners "provide data subjects with the types of information a reasonable data subject would want to know in a manner calculated to achieve understanding."[116]

To establish proper transparency, I propose that we take Pike's recommendation on transparency and expand its use. Not only should information about the logic and pathway of data and algorithms be provided to data subjects, but also to employees and data scientists. In other words, transparency about data and algorithm use should be common practice both internal and external to an organisation. The Childcare benefits affair exhibits the dangers of a black box model: it became difficult for employees to criticise the model and its outcomes. Transparency towards applicants

[111] Muller, "The Impact of AI on Human Rights," 14-15.

[112] Pike, "Defending Data," 740.

[113] Ibid.

[114] Ibid.

[115] Ibid.

[116] Ibid.

lacked as well, which made it impossible for them to properly raise concerns as well. Therefore, organisations should make understandable information available to both its employees and to its data subjects.

Explainability has to accompany transparency. Transparent information about an algorithm can only be provided if it is at least partially understood. Explainability techniques such as LIME (Local Interpretable Model-Agnostic Explanations), SHAP (SHapley Additive exPlanations) or DeepLIFT (Deep Learning Important FeaTures) can be utilised to establish partial explainability.[117]

For example, DeepLIFT provides explainability by determining the importance of model features. First, it determines a reference input, which is an uninformative background value.[118] In case of image classification the reference input is a completely black image. The activity of each neuron in the network is determined for the reference input.[119] Next several inputs, each containing a singular feature are fed into the model.[120] In case of image classification, different parts of the image are fed into the model. If the task is for example to recognise dogs or cats, the faces, tails, legs, and bodies are fed separately into the model. For each feature, the difference in neuron activation with the reference input is determined.[121] The bigger the difference, the higher the importance of the feature is estimated.[122] Although explainability techniques such as DeepLIFT illuminate the workings of the model, they do not generate full explainability. Therefore, further research into the explainability of self-learning algorithms is vital.

4.3 Governance

To properly accommodate ethical data and algorithm use, organisations should adjust their governance to integrate appropriate systems. I mention three possible governance-related solutions and refer to cited literature for more solutions. Firstly, appropriate governance only succeeds with proper guidelines. An ethical framework provides such a clear guideline. Although ethical frameworks are never all-

[117] John McDermid et al., "Artificial intelligence explainability: the technical and ethical dimensions," *Philosophical Transactions of the Royal Society* 379 (August, 2021): 8-9, https://doi.org/10.1098/rsta.2020.0363.

[118] Ibid.

[119] Ibid.

[120] Ibid.

[121] Ibid.

[122] Ibid.

encompassing, they establish awareness, common values and baseline expectations.[123] Secondly, responsibility should be clearly allocated within an organisation. Positions such as ethics officers can safeguard ethical data and algorithm usage.[124] Finally, an internal or external review board can be installed to perform ethical assessments of data and algorithms in which all stakeholders are considered.[125] Ethics officers could bring problematic algorithms to the attention of the review board.

4.4 Regulation

Proper regulation obligates ethical data and algorithm usage. Such regulation about ethical data and algorithm usage should come from an international level, as explained in Section 1.3. At this juncture, relevant international regulation is being developed both by the United States and the European Union.[126] For the orientation of this thesis, I focus on the European Union (EU) AI Act. This act will influence data and algorithm usage world wide, due to its extraterritorial application and demonstration effect.[127] Extraterritorial application of the EU AI Act flows from non-EU organisations that provide products or services to countries within the EU.[128]

The European Commission (EC) first submitted a proposal for the EU AI act in April 2021. The guidelines and restrictions in the act are differentiated by a risk scale, ranging from minimal to unacceptable risk.[129] Before the proposal becomes final legislation, formal procedures to reach consensus need to be followed.[130] Amendments have been proposed as recent as May 13, by the French Presidency of the Council, regarding regulation of general purpose AI systems.[131] A joint report of amendments is expected in fall 2022.[132] At this juncture, disagreement prevails over the definition of AI,

---

123 Pike, "Defending Data," 740; Muller, "The Impact of AI on Human Rights," 15.

124 Pike, "Defending Data," 741.

125 Pike, "Defending Data," 741; Muller, "The Impact of AI on Human Rights," 15.

126 Josh Meltzer and Aaron Tielemans, "The European Union AI Act: Next steps and issues for building international cooperation," *Brookings* (May 2022): 1, https://www.brookings.edu/wp-content/uploads/2022/05/FCAI-Policy-Brief_Final_060122.pdf.

127 Ibid.

128 Jos de Mul, personal communication, 3 June 2022.

129 Meltzer and Tielemans, "The European Union AI Act," 3.

130 Ibid.,1.

131 French Presidency of the European Commission, "Laying Down Harmonised Rules on Artificial Intelligence and Amending Certain Union Legislative Acts," May 13 2022, https://artificialintelligenceact.eu/wp-content/uploads/2022/05/AIA-FRA-Art-34-13-May.pdf.

132 Meltzer, "The European Union AI Act," 2.

the risk levels and governance structure to implement the act.[133] Consensus should be reached regarding these subjects. In addition, I recommend that dataficational problems and solutions are considered before the finalisation of the act.

[133] Ibid., 3-5.

Conclusion

In conclusion, datafication constitutes a problematic ideology. I have constituted four main problems of datafication: (1) the image of data as objective, (2) bias, unexplainability and intransparency in data processing, (3) difficulties in control of data usage, and (4) dataveillance. The scandal of the Childcare benefits affair and more specifically its discriminatory and illegal use of the risk-classification-model demonstrates how these problems manifest in practice. Dataficational problems underly many of the reasons that lead up to the Childcare benefits affair. For example, intransparency about data processing turned the model into a black box, which removed critical safety nets. Difficulties in control lead to late constitution of the problems.

I have established four solutions to dataficational problems as manifested in the Childcare benefits affair. Firstly, awareness of the non-objectivity (or intersubjectivity) of data and the problems of data processing should be increased through training. In that manner, the general public becomes more critical of data and data driven decisions. Secondly, transparency about the use of data and algorithmic models is conditional for an increased critical attitude. Explainability should accompany transparency. One can only be truly transparent about something if one can explain it. Thirdly, to integrate ethical use of data and algorithms within an organisation, I have suggested several governance-related solutions. These solutions are the establishment of: an ethical framework, ethics officers, and an ethics board. Finally, regulation is essential to make ethical use of data and algorithms completely mandatory. The EU AI Act constitutes a critical move to the right regulation. Dataficational problems should be considered before finalising this regulation.

The implications of this thesis stem from the established solutions. Although further research may be required, this thesis implies the implementation of training about the non-objectivity (or intersubjectivity) of data and the problems of data processing in school programs. Further research and requirements should be established regarding explainability and transparency. Other implications are the integration of proposed governance solutions and the establishment of international regulation. That way, the devil might be removed from the data!

Acknowledgements

I would like to thank professor doctor Jos de Mul, for his extensive feedback and valuable insights. His constructive comments took this thesis to a higher level. Many thanks also to my second corrector professor doctor Muller, whose analytical comments challenged me to further sharpen my conceptual analyses. Lastly, I'd like to acknowledge everyone that proofread this thesis and provided it with valuable critique.

Bibliography

Amnesty International. "Xenofobe machines: Discriminatie door ongereguleerd gebruik van algoritmen in het Nederlandse toeslagenschandaal." London: Amnesty International Ltd, October 2021. https://www.amnesty.org/en/documents/eur35/4686/2021/nl/.

Andrejevic, M. "Exploitation in the data-mine." In *Internet and Surveillance: The Challenges of Web 2.0 and Social Media*, edited by Christian Fuchs, Kees Boersma, Anders Albrechtslund, and Marisol Sandoval, 71-88. New York: Routledge, 2012.

Autoriteit persoonsgegevens. "Belastingdienst/Toeslagen De verwerking van de nationaliteit van aanvragers van kinderopvangtoeslag." Rijksoverheid, July 7, 2020. https://www.autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/onderzoek_belastingdienst_kinderopvangtoeslag.pdf.

Autoriteit Persoonsgegevens. "Boete Belastingdienst voor zwarte lijst FSV." Published April 12, 2022. Accessed May 2, 2022 https://autoriteitpersoonsgegevens.nl/nl/nieuws/boete-belastingdienst-voor-zwarte-lijstfsv#:~:text=Bovendien%20kreeg%20de%20Belastingdienst%20in,keren%20in%20de%20fout%20gegaan.

Auditdienst Rijk. "Onderzoeksrapport Toeslaggerelateerde CAF-zaken." Rijksoverheid, March 12, 2020. https://open.overheid.nl/repository/ronl-1c359fdc-4398-44e2-9969-1f353c0c481d/1/pdf/rapport-caf-adr.pdf.

Baise, Arnold. "The objective–subjective dichotomy and its use in describing probability." *Interdisciplinary Science Reviews* 45, no. 2 (2020): 174-185. https://doi-org.eur.idm.oclc.org/10.1080/03080188.2019.1705559.

Boyd, Danah, and Kate Crawford. Critical questions for Big Data. *Information, Communication & Society* 5, no. 15 (May 2012): 662-679. https://doi.org/10.1080/1369118X.2012.678878.

Brivot, Marion and Yves Gendron. "Beyond panopticism: On the ramifications of surveillance in a contemporary professional setting." *Accounting, Organizations and Society* 36, no. 3 (April 2011): 135-155. https://doi.org/10.1016/j.aos.2011.03.003.

De Mul, Jos. "The Informatization of the worldview." In: *Information, Communication & Society* 2, no. 1 (1999), 69-94. https://doi.org/10.1080/136911899359763.

De Mul, Jos. "Hoe de overheid vertrouwen verspeelt - en kan heroveren." *Trouw*, Letter & Geest, (July 2015): 4-7. https://www.trouw.nl/nieuws/hoe-de-overheid-vertrouwen-verspeelt-en-kan-heroveren~bda40169/.

De Mul, Jos. "The Living Sign. Reading Noble from a Biosemiotic Perspective." *Biosemiotics* 14, (May, 2021): 107-113. https://doi-org.eur.idm.oclc.org/10.1007/s12304-021-09426-y.

De Mul, Jos. "Help ik ben een database!" *De Groene Amsterdammer*, March 13, 2022, 46-49. https://www.groene.nl/artikel/help-ik-ben-een-database.

De Mul, "Metaphors we nudge by." In *Nudging media*, edited by James Katz. London: Palgrame MacMillan, unpublished manuscript, 10 May 2022.

Eitel-Porter, Ray. "Beyond the promise: implementing ethical AI." *AI and Ethics* 1, (October, 2021): 73-80. https://doi.org/10.1007/s43681-020-00011-6.

French Presidency of the European Commission. "Laying Down Harmonised Rules on Artificial Intelligence and Amending Certain Union Legislative Acts." May 13 2022. https://artificialintelligenceact.eu/wp-content/uploads/2022/05/AIA-FRA-Art-34-13-May.pdf.

Gitelman Lisa. *'Raw Data' is an Oxymoron.* Cambridge: MIT Press, 2011.

Herderscheê, Gijs. "Ruim 1.100 kinderen van gedupeerden toeslagenaffaire werden uit huis geplaatst." *De Volkskrant*, October 19, 2021. *https://www.volkskrant.nl/nieuws-achtergrond/ruim-1-100-kinderen-van-gedupeerden-toeslagenaffaire-werden-uit-huis-geplaatst~baefb6ff/*.

Horkheimer, Max and Theodor Adorno. *Dialectic of Enlightenment.* Translated by Edmund Jephcott. Edited by Gunzelin Schmidt Noerr. Stanford: Stanford University Press, 2002.

Huffelen, A.C. van. "Kamerbrief inzake de openbaar making van het risicoclassificatiemodel toeslagen." *Rijksoverheid*, November 26, 2021. https://

open.overheid.nl/repository/ronl-aab59576-16e3-49ce-84f2-d57afc760bbb/1/pdf/kamerbrief-inzake-de-openbaar-making-van-het-risicoclassificatiemodel-toeslagen.pdf.

Kitchin, Rob and Gavin McArdle, "What Makes Big Data, Big Data? Exploring the Ontological Characteristics of 26 Datasets", *Big Data & Society* 3, no. 1 (2016). doi: 10.1177/2053951716631130

Mayer-Schönberger, Viktor and Kenneth Cukier. *Big Data: A Revolution That Will Transform How We Live, Work and Think*. London: John Murray, 2013.

McDermid, John, Yan Jia, Zoe Porter and Ibrahim Habli. "Artificial intelligence explainability: the technical and ethical dimensions." *Philosophical Transactions of the Royal Society* 379 (August 2021). https://doi.org/10.1098/rsta.2020.0363.

Mejias, Ulises, and Nick Couldry. "Datafication." *Internet Policy Review* 8, no. 4 (November 2019): 1-10. https://doi.org/10.14763/2019.4.1428..

Meltzer, Josh and Aaron Tielemans. "The European Union AI Act: Next steps and issues for building international cooperation." *Brookings* (May 2022). https://www.brookings.edu/wp-content/uploads/2022/05/FCAI-Policy-Brief_Final_060122.pdf.

Ministerie van Financiën. "Gebruik van dubbele nationaliteit." December 18, 2020. https://services.belastingdienst.nl/toeslagen-herstel/gebruik-van-dubbele-nationaliteit/.

Muller, Catelijne. "The Impact of Artificial Intelligence on Human Rights, Democracy and the Rule of Law: draft d.d. March 20, 2020." *ALLAI*, March 20 2020. https://allai.nl/wp-content/uploads/2020/06/The-Impact-of-AI-on-Human-Rights-Democracy-and-the-Rule-of-Law-draft.pdf.

Organisation for Economic Co-operation and Development. *OECD glossary of statistical terms*. Paris: OECD publishing, 2008.

Pike, Elizabeth. "Defending Data: Toward Ethical Protections and Comprehensive Data Governance." *Emory Law Journal* 69 (April, 2019), 687-742. https://scholarlycommons.law.emory.edu/elj/vol69/iss4/2.

Parlementaire ondervragingscommissie Kinderopvangtoeslag. "Tweede Kamer der Staten-Generaal, Verslag - Parlementaire Ondervragingscommissie

Kinderopvangtoeslag: Ongekend Onrecht." Tweede Kamer, December 17, 2020. https://www.tweedekamer.nl/sites/default/files/atoms/files/20201217_eindverslag_parlementaire_ondervragingscommissie_kinderopvangtoeslag.pdf.

Rijksoverheid, "Bijlage 16-26," November 26, 2021, https://www.rijksoverheid.nl/documenten/publicaties/2021/11/26/overige-bijlagen-bij-kamerbrief-over-openbaar-making-risicoclassificatiemodel-toeslagen.

Rijksoverheid. "Bijlage 69-75." November 26, 2021. https://www.rijksoverheid.nl/documenten/publicaties/2021/11/26/overige-bijlagen-bij-kamerbrief-over-openbaar-making-risicoclassificatiemodel-toeslagen.

Rijksoverheid. "Werkbeschrijving Risicoselectie." March 2014. https://www.rijksoverheid.nl/documenten/rapporten/2014/03/31/werkbeschrijving-risicoselectie---maart-2014.

Siau, Keng and Weiyu Wang. "Artificial intelligence (AI) ethics: ethics of AI and ethical AI." *Journal of Database Management* 31, no. 2 (April, 2020): 74-87. https://doi.org/10.4018/JDM.2020040105.

Von Esbach, Warren. "Transparency and the Black Box Problem: Why We Do Not Trust AI." *Philosophy & Technology* 34 (September 2021): 1607-1622. https://doi.org/10.1007/s13347-021-00477-0

Van Dijck, José. "Datafication, Dataism and Dataveillance: Big Data Between Scientific Paradigm and Ideology." *Surveillance & Society* 12, no. 2 (May 2014): 197–208. https://doi.org/10.24908/ss.v12i2.4776.