Designing for Responsibility

Sattlegger, Antonia; Van Den Hoven, Jeroen; Bharosa, Nitesh

DOI
10.1145/3543434.3543581

Publication date
2022

Document Version
Final published version

Published in
Proceedings of the 23rd Annual International Conference on Digital Government Research

Designing for Responsibility

Antonia Sattlegger
Delft University of Technology
a.s.sattlegger@tudelft.nl

Nitesh Bharosa
Delft University of Technology
n.bharosa@tudelft.nl

Jeroen van den Hoven
Delft University of Technology
m.j.vandenhoven@tudelft.nl

ABSTRACT

Governments are increasingly using sophisticated self-learning algorithms to automate and standardize decision-making on a large scale. However, despite aspirations for predictive data and more efficient decision-making, the introduction of artificial intelligence (AI) also gives rise to risks and creates a potential for harm. The attribution of responsibility to individuals for the harm caused by these novel socio-technical decision-making systems is epistemically and normatively challenging. The conditions necessary for individuals to be adequately held responsible – moral agency, freedom, control, and knowledge, can be undermined by the introduction of algorithmic decision-making. Thereby responsibility gaps are created where seemingly no one is sufficiently responsible for the system's outcome. We turn this challenge to adequately attribute responsibility into a design challenge to design for these responsibility conditions. Drawing on philosophical responsibility literature, we develop a conceptual framework to scrutinize the task responsibilities of actors involved in the (re-)design and application of algorithmic decision-making systems. This framework is applied to an empirical case study involving AI in automated governmental decision-making. We find that the framework enables the critical assessment of a socio-technical system's design for responsibility and provides valuable insights to prevent future harm. The article addresses the current academic and empirical lack of philosophical insights to understand and design for responsibilities in novel algorithmic ICT systems.

KEYWORDS

AI, Digital Government, Algorithmic Decision-Making, Task Responsibility

ACM Reference Format:
Antonia Sattlegger, Nitesh Bharosa, and Jeroen van den Hoven. 2022. Designing for Responsibility. In *DG.O 2022: The 23rd Annual International Conference on Digital Government Research (dg.o 2022), June 15–17, 2022, Virtual Event, Republic of Korea.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3543434.3543581

1 INTRODUCTION

Governments are increasingly emphasizing the potential of artificial intelligence (AI) in the public sector [1]. AI promises more efficient, data-driven, and evidence-based public administration [2]. AI is applied across different public domains, including the automation of decision-making and service provision in areas critical to the well-being of vulnerable citizens, such as welfare services, healthcare provision, or policing [3–5]. Beyond the promise of more efficient, data-driven, and evidence-based public administration, there is a darker side to the application of AI in the public sector [4]. The application of AI in the 'digital welfare state' is particularly concerned. In his report on digital welfare states and human rights, UN Special Rapporteur Philip Alston [6] argues that governments increasingly use "predictive analytics to foresee risk, automate decision-making, and remove discretion from human decision-makers" [6] in their obsession with "fraud, cost savings, sanctions, and market-driven definitions of efficiency" [6]. This digital transformation "disproportionally targets the poorest and most marginalized in society" [6].

This dark side of AI was illustrated strikingly in the Dutch childcare benefits scandal. The scandal surrounding a nationwide applied algorithmic fraud indication system has emphasized AI systems' potential risks and harms in administrative decision-making. One may argue that the Dutch childcare benefits scandal was the first case in which a government fell over the irresponsible application of AI. A nationwide applied self-learning algorithmic risk classification system was applied for the automated screening and predictive assessment of childcare benefits applications on a greater scale. Eventually, it was found that the self-learning algorithmic system assessed applications for childcare benefits based on several dozen indicators, including the distance between the residence and the childcare facility, as well as discriminatory indicators, such as income and citizenship [7]. The discriminatory and erroneous assessment led to stigmatizing and unfounded fraud investigations. Families were unjustly forced to pay back tens of thousands of euros over minute errors, such as missing signatures, with no means of redress. Children were removed from their homes. Families and livelihoods were ruined. Consequently, 71% of the public say their trust in the political system was negatively impacted [8].

The Dutch childcare benefits scandal illustrates the risks of applying AI in government decision-making. One of the core challenges is to define who is responsible for the harm caused and, perhaps even more important, how to prevent harm in the future. The use of algorithms in socio-technical decision-making systems requires us to rethink the allocation of responsibilities and the conditions that need to be fulfilled for individuals to take these responsibilities. We argue that the application of AI in governmental decision-making may challenge the adequate individual attribution of responsibility for the outcomes of such systems. Philosophers generally distinguish three conditions necessary for the adequate attribution of responsibility to individuals. Firstly, the individual must be able understand the moral significance of her actions and act accordingly. Secondly, the individual must be able to act freely and without coercion. Thirdly, the individual must possess sufficient knowledge to be aware of the consequences of one's actions. However, these

dg.o 2022, June 15–17, 2022, Virtual Event, Republic of Korea

ACM ISBN 978-1-4503-9749-0/22/06.
https://doi.org/10.1145/3543434.3543581

conditions can be undermined for both designers and operators due to the properties of AI systems. AI is understood as autonomous, interactive, and adaptive technology that is capable to carry-out tasks which require human-like intelligence [36]. Several design properties of AI in socio-technical decision-making systems challenge the attribution of moral responsibility [37]. Firstly, it's learning capabilities enable the system to evolve in the interaction with its environment, making it unpredictable over time. Secondly, AI systems are often designed as black boxes. Opaque AI decision-making processes may be difficult to explain and predict for both designers and users of the system. Thirdly, AI may be designed with varying degrees of autonomous decision-making capabilities, further challenging meaningful human control over the system. Lastly, many stakeholders are involved in and affected by the (re-)design, development, and application of AI systems. These stakeholders may have different capacities and preferences in the interaction with the system [37]. The involvement of 'many hands' [27] challenges the attribution of responsibility to individual actors.

We turn this challenge to adequately attribute responsibility into a design challenge. Rather than looking back and attributing responsibility after the fact, we argue that these responsibility conditions need to guide the design of socio-technical decision-making systems applying AI to prevent future harm. Our main research questions are threefold:

First, which responsibilities should be attributed to relevant individuals when applying AI in governmental decision-making? Second, which conditions need to be satisfied for the adequate attribution of these responsibilities? How do we design for these conditions?

This paper proceeds as follows: Section two discusses the research method. Section three presents the conceptual framework for moral responsibility and the respective conditions to attribute individual moral responsibility fairly and effectively. Section four presents the case study analysis. Lastly, we will discuss how the research of past harm can contribute to preventing future harm. We shall propose theoretical insights and practical recommendations for the responsible design and application of AI decision-making systems in the public sector.

2 METHODOLOGY

We first develop a conceptual framework on moral responsibility that we apply in a single case study. The case of the algorithmic risk classification model was chosen as a suitable case [9] based on the following criteria. First, the application of the algorithmic model was found to have contributed to significant harm to those deemed to be fraudulent in the application for childcare benefits. Eventually, the government coalition and prime minister stepped down over the unfolding scandal. Second, there is much data available on the case. Due to the public interest, detailed documentation about the algorithmic model's design, use, and impact have been publicized. The case has been thoroughly reviewed and evaluated by national and international government and independent, non-governmental organizations. The data collection was a systematic search for governmental reports, independent inquiries, and subsequent snowballing technique. The empirical analysis was based on document analysis of government reports and non-governmental inquiries into the child benefits scandal. The documents were analyzed using deductive coding within the NVivo software [9]. The conceptual framework developed below served as the coding book.

3 CONCEPTUAL FRAMEWORK

In this section, we define the key concepts of this research, which include (1) AI in public sector decision-making and (2) the philosophical foundation of moral responsibility. Subsequentially, we develop a conceptual framework by drawing on the notions of what has been called "task-responsibility" [21] and "meta-task responsibility" [23] in connection with the necessary conditions that need to be in place for individuals to be held responsible.

3.1 Artificial Intelligence in the Public Sector

As the number of studies on AI in the public sector is growing [e.g. 5, 10], we are steadily beginning to see the benefits and dangers more clearly. In the absence of generally accepted and unequivocal terminology, it is essential to clarify the application of these terms and their underlying assumptions. Algorithmic systems may carry out tasks that require intelligent, human-like behavior. The transformative difference that AI makes in the public sector is not merely the digitalization of traditional administrative decision-making procedures but a new quality of predictive analytics and autonomous decision-making. It is essential to consider the application of algorithms as part of a broader socio-technical system. Isolated technical or legal solutions to the responsible design of artifacts, such as explainable, transparent, or responsible AI, have largely neglected this embedding – an optimism Stilgoe [11] refers to as "technical solutionism". Kitchin [12] emphasizes that "algorithms need to be understood as relational, contingent, contextual in nature, framed within the wider context of their socio-technical assemblage. From this perspective, 'algorithm' is one element in a broader apparatus which means it can never be understood as a technical, objective, impartial form of knowledge or mode of operation." [12]. Selbst et al. [13] problematize the lack of considering AI as part of broader socio-technical systems in current (computer science) discussions on just and fairness-aware learning algorithms. A failure to understand the interactions between the technical systems and social worlds leads to two different traps: the framing trap – a "failure to model the entire system over which a social criterion, such as fairness, will be enforced" [13], and the formalism trap – a "failure to account for the full meaning of social concepts such as fairness, which can be procedural, contextual, and contestable, and cannot be resolved through mathematical formalisms" [13]. The authors [13] argue that, for one, "technical designers can mitigate the traps through a refocusing of design in terms of process rather than solutions" (p. 59) and for another, must "include social actors rather than purely technical ones." [13]. This research addresses these traps and focuses on algorithmic decision-making systems as socio-technical systems.

3.2 Individual Moral Responsibility

Moral Responsibility is not a single, unitary, and generic concept [14], but a polysemous term that admits a variety of meanings, usages, and degrees. A variety of different taxonomies of moral responsibility have been developed to illustrate the scope of moral

responsibility [see earliest 15; or more recent 14, 16, 17]. In the philosophical literature being morally responsible is often understood to imply, among other things, that "the person is an appropriate candidate for reactive attitudes" [18]. These reactive attitudes encompass a spectrum of feelings directed at a responsible agent for her action or contribution to an outcome. They can be positive reactions, feelings, and attitudes, such as praise, gratitude, or respect, or negative reactions such as blame and resentment [18, 19].

This backward-looking notion of moral responsibility has dominated much of the traditional philosophical literature. In this sense, moral responsibility has been primarily understood as backward-looking responsibilities and referring to past actions to enable the normative evaluation, such as consequential blame or praise (responsibility as blameworthiness) and possible retribution (responsibility as liability) [20]. Another backward-looking notion of responsibility that has been prevalent in public administration studies is responsibility-as-accountability. An accountable agent has a responsibility to provide an account of her actions and why she is not blameworthy for a state-of-affairs. While these backward-looking responsibilities can have a behavioral effect in stimulating desirable or discouraging undesirable actions, the actors are made responsible for past outcomes. These responsibilities are attributed ex-post to individuals based on their contribution to the harm that has already taken place. However, we are interested in the prevention of harm through the ex-ante design for responsibility. Thus, we focus on forward-looking responsibilities which prescribe responsibility to individuals for future state-of-affairs [16]. Two kinds of forward-looking notions of responsibility can be distinguished: Firstly, responsibility-as-virtue attributes responsibility to a person rather than a state of affairs. A responsible agent is characterized by the willingness to take responsibility and acts in due care for others. Secondly, responsibility-as-obligation implies that one has the (moral) obligation to see to it that a certain state-of-affairs is brought about. Goodin [22] formulated this obligation as follows: "i ought to see to it that φ." [22]. This formulation specifies that the obligation does not refer to specific actions on behalf of the responsible actor, such as "i does or refrains from doing α" [22] but bringing about a desirable state-of-affairs. Likewise, it is insufficient that the desired state-of-affairs (φ) occurs, the core of the obligation lies in 'seeing to it'. The obligation – "ought to" – can stem from multiple sources, such as legal, cultural, or normative assumptions. We can think of 'ought' as having an index or subscript 'ought, following Dutch criminal law,' 'ought, following our contract,' 'ought, following the declaration of human rights, 'ought, according to the moral point of view'. It is important to emphasize the source of the responsibility as a moral obligation rather than the descriptive task attribution. Van de Poel [16] illustrates this with a striking example: "Whereas it might be said that Eichmann had the task (responsibility) that the Jews were effectively transported to the concentration camps, it does not follow that he had a (moral) obligation to see to it that they were effectively transported. In fact, because the transport was part of an immoral plan, aiming at the extinction of the Jews, he might even have had a moral obligation to see to it that they were not effectively transported." [16]. Responsibility-as-obligation is particularly relevant regarding the design for responsibility. This notion of responsibility enables the fine-grained organization of responsibility by ex-ante specifying abstract desirable state-of-affairs into concrete individual obligations. This specification is difficult in complex and uncertain innovation processes in which new normative requirements may arise throughout the process, especially when these responsibilities are attributed externally [17]. However, it is precisely for this difficulty that responsibility(as-obligation) needs to be part of the design process, rather than ex-post when individuals refuse to take responsibility for harm that has already occurred.

Goodin's [22] conceptualization of responsibility-as-obligation into task-responsibilities enables this specification. He further specifies the obligation 'to see to it that' into "activities of a self-supervisory nature" [22]. These "require minimally: that i satisfy himself that there is some process (mechanism or activity) at work whereby X will be brought about; that i check from time to time to make sure that that process is still at work and is performing as expected; and that i take steps as necessary to alter or replace processes that no longer seem likely to bring about X.". The different types of responsibility-as-obligation can be deduced. They are summarized in Table 1.

Van den Hoven [23] applies Goodin's [22] concept of task-responsibility to elucidate the responsibilities of designers and users of decision support systems in the public sector. Van den Hoven [24] has referred to such environments as artificial epistemic niches. As decision-makers, the civil servant can become 'narrowly embedded' in these digital government systems because the system presents itself as a black box to the end-user. The operator is epistemically dependent upon the system because he has to defer to it for the justification of his actionable beliefs, he lacks the independent epistemic resources to contest the output of the system because he is dependent on the knowledge the system produces and the logic it uses and is not able to scrutinize processes during run time. This is applied to complex rule-based systems, but it applies a fortiori to AI systems. The operator cannot put forward "system independent reasons" for her beliefs or for reasons to overrule or disagree with the system [23]. The operator cannot morally justify deviating from the system at the moment of decision-making. This argumentation pointing to the moral consequences of epistemic dependence and narrowly embedded end-users is very relevant to understanding the moral predicament of operators and screen-level bureaucrats in the age of ubiquitous AI applications in public administration. In order to take their task-responsibilities, civil servants as end-users of the decision support systems "(. . .) ought to endorse (or act upon) the output of Information Systems they are epistemically dependent upon, and with which they know they will be working under conditions of narrow embeddedness, only after an inquiry of acceptability of the system, the cost of which is proportional to the cost that could reasonably be expected if what is endorsed and acted upon should prove in any sense to be inadequate." (p. 106). To morally empower the civil servant in the use of AI systems, designers "ought to allow users to work with systems in such a way as not to make it impossible for them to live up to their obligations as users." (p. 106). They have thus a meta-task responsibility to design "the system or epistemic artifact (. . .) in such a way as to allow the user to work with it, while retaining his status as a morally autonomous person, who can take his responsibility." (p. 106). If a task-responsibility of agent A for X is an obligation to see to it that X is carried out, then there is a meta-task responsibility associated

Table 1: Task responsibilities – Which task-responsibilities can be attributed to individuals based on their obligations?

Task responsibilities – Which task-responsibilities can be attributed to individuals based on their obligations?	
Task responsibility	Task responsibility implies the obligation to see to it that X is brough about [22].
Negative task responsibility	Negative task responsibility regarding X implies the obligation to see to it that no harm is done in seeing to it that X is brought about [22].
Self-monitoring responsibility	Self-monitoring responsibility implies that A ought to "satisfy himself that there is some process (mechanism or activity) at work whereby X will be brought about; that A check, from time to time, to make sure that that process is still at work and is performing as expected; and that A take steps as necessary to alter or replace processes that no longer seem likely to bring about X." [22].
Supervisory responsibility	Supervisory responsibility implies that A ought "to see to it that others act or refrain from acting in a certain way" [22].
Meta-task responsibility	"A has an obligation to see to it that (1) conditions are such that it is possible to see to it that X is brought about, (2) conditions (moral agency, knowledge, freedom, capacity) are such that it is possible to see to it that no harm is done in seeing to it that X is brought about." (. . .) while retaining her status as a morally autonomous person, who can take moral responsibility [23].

with it: relevant others and A himself have an obligation to see to it that A can see to it that X is done. If A has a task-responsibility to make the correct payments on Friday, then A – and relevant others – have on obligation to see to it before Friday (as far as this is possible) that A can do what A has to do on Friday, and they have an obligation to refrain from doing such things that prevent A from doing so (p. 106). Meta-task responsibility plays an essential role in effectively distributing (task-) responsibilities across the multi-actor design and usage of complex algorithmic decision-making systems. Building on Goodin's [22] expressions of task-responsibility, Van den Hoven [23] defines meta-task responsibility as follows: "A has an obligation to see to it that (1) conditions are such that it is possible to see to it that X is brought about, (2) conditions are such that it is possible to see to it that no harm is done in seeing to it that X is brought about." (p. 108). Table 1 summarizes these task- and meta-task responsibilities.

3.3 Conditions for individual moral responsibility

To adequately ascribe responsibility to a person for an action, certain conditions need to be fulfilled. These responsibility conditions are also described as "fairness criterion of responsibility ascriptions" [24]. These preconditions enable an individual to take responsibility for fulfilling her obligations. What these conditions amount to can be seen by studying generally accepted types of excuses and viable attempts to deny one's responsibility. If there is damage to a precious object, for example, and someone is held responsible for the untoward outcome, we often hear: "I was not the one who caused it", "there is nothing wrong with it", "it was not my intention", "I was forced, I had no choice" or "I did not know what was happening". Excuses target the conditions for responsibility, such as intention, knowledge, capacity to judge, free choice, causal involvement, something that went wrong. In the case of black box AI-based systems, it is evident that many of these excuses are readily available. Plausible deniability of responsibility is almost guaranteed if users are in the dark. Rubel, Castro & Pham [25] emphasize that users and designers of algorithmic decision-making systems, can intentionally obscure their moral responsibility – thus, engaging in agency laundering. According to the authors "using an automated process to make decisions can allow a person to distance herself from morally suspect actions by attributing the decision to the system" [25] and "letting it forestall others from demanding an account for bad outcomes that result" [25], thereby laundering their agency.

There is no universal agreement regarding the formulation of these conditions [26]. Multiple typologies have been developed to distinguish these conditions [see e.g. 18, 19, 20, 27, 28]. Generally, three interrelated conditions are necessary for the fair and effective attribution of moral responsibility can be distilled, as summarized in Table 2

- Moral agency and intentionality – Moral responsibility presupposes a moral agent capable of intentional and purposeful action. She can grasp moral reason and can control her behavior accordingly. An agent's action or inaction stems from a decisional mechanism responsive to moral reason. This mechanism is receptive to (moral) reasons for and against a particular course of action, as well as reactive to those (moral) reasons [18]. This condition is violated if an agent acts under force or the influence of drugs.
- Freedom and control – Interrelated with the condition of moral agency is the condition of free will or control over one's action. A responsible person must be capable of determining and acting according to one's moral reasoning. Free will requires the absence of coercion, force, or other barriers outside the actor's control. The agent must take ownership of one's decisional mechanisms to be able to take moral responsibility. Thus, if one decides to apply an algorithmic system for sensitive decision-making, one cannot blame the algorithmic system for its output as one has taken ownership of the decisional mechanism in the first place. Thus, the agent's action "issues from the agent's own, reason-responsive mechanism." [18]. The meaning and scope of free will continue to be disputed. However, few would dispute that an actor who acts under coercion can be morally responsible for her actions.

Table 2: Responsibility conditions – When is it adequate to attribute responsibility to someone?

Responsibility conditions – When is it adequate to attribute responsibility to someone?	
Moral agency & intentionality	A responsible actor can engage in intentional, purposeful action. She understands the moral significance of her action and can reason accordingly.
Freedom & control	A responsible actor can act freely and without coercion. The actor has control and can take ownership over the decisional reason-responsive mechanisms.
Knowledge	A responsible actor possesses sufficient knowledge to be aware of the consequences and causal contributions of one's action or inaction.

- Knowledge – Knowledge and awareness are essential epistemic conditions for moral responsibility. A responsible actor is aware of the consequences and causal contributions of one's action or inaction. An actor is not excused for ignorance due to negligence. An actor has the normative duty to ensure she knows what she should know or can reasonably be expected to know.

These responsibility conditions enable a systematic assessment of an agent's capacity for being attributed moral responsibility. This does not mean that these actors are entirely excused if these conditions are not fulfilled. An obvious case, for example, is when their ignorance is self-caused since self-caused ignorance (what is called ignorantia affectata) does not excuse. Likewise, epistemic recklessness does not excuse. In what follows, we assume that only human beings can be morally responsible. The conditions of moral agency and freedom are closely interrelated and form inherently human conditions for being responsible and being able to take responsibility. In a backward-looking sense of responsibility, both human agents and non-human agents, such as artificial agents, or even natural events, such as the weather, can be causally responsible for an event, but only human agents can be morally responsible [18]. While the system crash of one's computer can be causally responsible for wiping out weeks' worth of work, we intuitively feel silly to react with emotional resentment or blame towards our computer. In the case of an artificial and non-human agent, such as our computer or an algorithmic model, the conditions necessary for the attribution of moral responsibility do not apply. For this argumentation, we will assume in this work that only human agents can – given these conditions – be moral agents and, thus, be morally responsible. This argumentation extends to collectives of individuals, such as government organizations or administrative units, who cannot be morally responsible. Eventually, those representing the collective are crucially involved in its decision-making and governance may be held morally responsible, not as private individuals but based on their professional function or role, such as public officials [27].

4 CASE STUDY – THE BENEFITS MACHINE: RESPONSIBILITY IN THE ALGORITHMIC SOCIO-TECHNICAL DECISION-MAKING SYSTEM

We apply the concepts of task- and meta-task responsibility discussed above to better understand how the application of the algorithmic risk classification model used by the Dutch Tax Authorities from 2013 until 2019 can cause severe harm and great injustices to Dutch citizens. Thereby, we seek to apply insights to the design for the responsibility of AI-based socio-technical systems to prevent future harm. We will analyze which responsibilities and conditions were attributed to the relevant actors tasked with designing, developing, and applying the algorithmic model. See Table 3 for an overview of the relevant actors.

The application of the algorithmic model must be understood as part of a socio-technical system that is embedded within a broader political and institutional context. Table 4 summarizes the relevant actors.

4.1 Responsibilities in designing the algorithmic risk classification model

The algorithmic model was designed by civil servants and data specialists from the department Allowances, a sub-department with the Dutch tax authorities, subordinate to the Ministry of Finance. The design choices made reflect the task responsibilities the Department was assigned within the government organization and attributed within this hierarchical order.

The Allowances department has the core task responsibility for granting, paying, and recovering childcare benefits. With a strong political priority on the efficient prevention and combat of fraud in the childcare benefits system, civil servants at the Allowances department were to check the applications before payments were made, particularly those without prior residence in the Netherlands. To do so more time and cost-efficiently, the Allowances department was tasked with developing an ICT system that could screen the citizen-clients for fraud automatically at scale. Table 5 (Appendix) summarizes the developer's task responsibilities.

The developers at the Allowances department of the tax authorities acted within a political and institutional environment that prioritized efficient prevention and combatting of fraud over other public values, such as the rule of law, such as "foreseeability, for those affected, precision and scope in the executive's discretion, and respect for human rights" [30] to prevent hardship for citizen-clients. Ministers argued that "the Tax and Customs Administration/Allowances should operate like a machine (. . .). Exceptions would throw a spanner in the works." [31]. These expectations were reflected in the General Act on Means-tested Benefits, the policy the tax authorities were expected to execute. Following this political prioritization, the Tax Authorities were pressured by the Ministry of Social Affairs to finance their enforcement scheme through the repayments and fines by the citizen-clients. Simultaneously, the Tax Authorities and sub-units were faced with significant spending

Table 3: The relevant actors and their roles in the design, development, and application of the algorithmic model

Relevant actors and their roles in the design, development, and application in the algorithmic decision-making system
Design of the algorithmic model
• Civil servants and data specialists from the Allowances department as designers of the algorithmic model
Application of the algorithmic model
• Civil servants as operators of the algorithmic model
• Citizen-clients as objects or targets of the decision-making process

Table 4: The relevant actors and their roles in the political and institutional context

Relevant actors and their roles in the design of the socio-technical decision-making system
Design of the social benefits legislation
• Government and Parliament
Design of the execution of the social benefits legislation
• Ministry of Social Affairs
• Ministry of Finance
Enforcement of the social benefits system
• Tax and Customs Administration

cuts imposed by the Cabinet, putting additional pressure on the streamlining and greater efficiency of the "benefits 'machine'". One may argue that this financial pressure led to an immoral incentive to maximize repayments and undermined the Tax Authorities' decisional reason-response mechanism. Professional discretion and room for free choice were deliberately designed out across all levels and tiers of government. As the former director of Allowances emphasizes, "the current benefits system is very complex (. . .). There is hardly any room for maneuver in administering it." [31]. This political and organizational background sheds light on the corrosion of the responsibility conditions that the system developers faced.

The self-learning algorithmic risk classification model was developed in 2013 by civil servants and data specialists from the department Allowances. A set of weighted risk indicators was selected based on a statistical analysis of historical data. A scorecard was developed based on which all incoming applications were automatically assessed, and a respective fraud risk score was calculated.

The algorithmic model automatically assessed monthly incoming applications for childcare benefits based on several dozen weighted indicators. Indicators related to the childcare center, such as the type of childcare or the distance between childcare center and residence, as well as to the situation of the applicant, such as income, benefit debts, family status, age, and the number of children [32]. Based on the accumulated risk score of all indicators, those applications were selected for subsequent manual scrutiny that were either above a risk score of 0.8 or part of 30-100 applications (depending on capacities) within a month that had the relatively highest risk score below 0.8. Civil servants then manually assessed those high-risk cases and had the power to label them as fraudulent. All other applications were automatically approved.[1]

In doing so, the developers at the Allowances department arguably had a so-called negative task responsibility to develop the

[1] Automated or semi-automated system – The algorithmic decision-making system can be understood as both an automated and semi-automated system. The algorithmic decision-making was fully automated for those that were not perceived to be of high risk by the system. Those with a high-risk score were automatically selected for manual scrutiny, thus, a semi-automated system with human interference. Due to this differentiation, it was not legally necessary to inform the citizen-clients about using an algorithm according to the GDPR.

system without (i) inflicting harm on innocent, non-fraudulent citizen-clients, (ii) placing unproportioned burdens on the end-user, and (iii) imposing additional costs on the administration, but arguably equally without making it impossible for end-users to take and bear moral responsibility for decisions significantly affecting citizens. The developers violated this crucial responsibility. Nationality was deliberately included as a risk indicator in the initial design of the algorithmic model. Not having the Dutch nationality lead to an increase in an applicant's risk score [7]. The algorithmic risk profiles meant that those of non-Dutch nationality or low income were disproportionally selected and, thus, affected by delays in benefits payments and the harsh sanctions that could follow from a manual assessment. Amnesty International [33] has criticized the algorithmic model for enabling intersectional discrimination and human rights abuse. Amnesty International [33] concluded that "the inclusion of nationality as a factor in the risk classification model, in combination with the mapping of certain groups of people whom the tax authorities believed would be more likely to engage in fraudulent or criminal behavior, shows that the tax authorities were motivated by racial prejudice concerning fraud detection. This risk-scoring led to a disproportionate focus on particular groups of people based on their ethnicity and qualified as racial profiling under the international human rights framework." [33]. There was no differentiation within the algorithmic model between different non-Dutch nationalities. However, in an exemplary case, the risk indication of 120 to 150 citizen-clients of Ghanaian nationality led to a subsequent manual investigation of all 6047 citizen-clients with Ghanaian nationality [7]. The model itself dropped the indicator nationality in October 2018 due to its low risk predicting power. It is assumed that publishing information in different languages on the tax authorities' website had led to fewer incorrect applications [7].

The developers had the self-monitoring responsibility to monitor and, if necessary, re-design the self-learning algorithmic decision-making system. This includes the responsibility to design the algorithmic model to allow for continued oversight and an understanding of its internal workings. This responsibility was neglected in the design of the algorithmic model as self-learning. The algorithmic model was a self-learning model that was continuously trained with historical data. The model continuously updated the weighted risk indicators based on the high-risk cases that were manually assessed and either deemed correct or incorrect assessments. Specific indicators were added others dropped throughout the lifespan of the model. This self-learning characteristic meant that the indicators and their weights were designed and re-design throughout the application by the algorithm itself without sufficient human understanding and oversight. This self-learning design meant that a sufficient understanding of the changing algorithm, continuous oversight, and deliberate re-design were lacking. A necessary consequence is that even though specific protected characteristics may (legally) not be collected or added as indicators (such as ethnicity), the algorithm can develop proxy variables (such as postal code), which can correlate strongly with the former. Eventually, those citizen-clients deemed to have the highest risk and subsequently scrutinized lived in an urban area, with an income under 20.000 Euro, were single parents, with multiple children in the household, who asked for many childcare hours, and who lived far from the childcare center [34].

Management had the supervisory responsibility to see to it that the developers saw to it that end-users of the model use the risk score as an indication for manual assessment could see to it that assessment of the applications could be carried out independently. However, there is no indication that the algorithmic model was evaluated and subsequentially re-designed [35].

Core responsibility is the meta-task responsibility of the developers and higher management, according to which they have the obligation to see to it that the end-user can use the algorithmic model responsibly. By designing for the end-user's responsibility conditions, the end-user is empowered to take responsibility for the outcomes of her application of the algorithmic model. Table 6 illustrates the end-users necessary responsibility conditions. These are the conditions and the non-functional requirements that the developers ought to design for.

As elaborated above, three interrelated responsibility conditions that enable civil servants to fulfill their respective task responsibilities can be distinguished. First, moral agency and intentionality, the end-user must work with the system as a morally autonomous person (within the bounds of the professional role) and understand her model-independent actions. Second, freedom & control, the end-user must be free in applying the algorithmic model. Third knowledge, the end-user should understand why and how the algorithmic model arrives at a certain risk indication. These conditions were undermined by the design of the algorithmic model and the way it was embedded in the organization. Firstly, the algorithmic model was autonomous insofar as the end-user is automatically tasked to scrutinize the applicant flagged as high risk by the model. She has no professional discretion or freedom on whether or not to proceed with the assessment. Secondly, the end-user was epistemically dependent on the risk score provided by the algorithmic model. She blindly relies on this information because she cannot scrutinize the processes during run time as the algorithmic model was designed as a black-box model. The end-user was not provided with additional information beyond the accumulated risk score [30]. She did not understand how and why an individual application was flagged with a high-risk status. She was not provided with further knowledge about the input and workings of the algorithmic model. Therefore, the end-user was epistemically dependent on the algorithmic model. Without means to scrutinize the model, she cannot justify her beliefs that an applicant is potentially fraudulent independent of the algorithmic model's output. Her manual judgment of the cases selected by the model can be intentionally or unintentionally impacted by the algorithm's risk indication. These conditions – narrow embeddedness in an artificial epistemic niche and epistemic dependency – result in the epistemic enslavement of the end-user [23]. In such circumstances, Van den Hoven [23] argues that "in order to curb the reduction of intellectual autonomy and relativization of moral responsibility, the user must be permitted to reflect ex-ante upon the epistemic conditions, within the confines of which she knows she will be working." (p. 106). There is no documentation that such ex-ante reflection was enabled either through deliberative processes or inclusive design practices. Instead, there are strong indications that end-users did not understand the workings of the black-box model [33].

4.2 The civil servant's task responsibilities in the application of the algorithmic model

The algorithmic model was designed based on two subsequent task responsibilities. Firstly, civil servants were expected to independently assess the application selected for manual scrutiny by the model. Secondly, the citizen-clients were expected to provide the correct information in their applications.

As the end-user of the algorithmic model, the civil servant has the task responsibility to independently assess those applications that are flagged as high-risk by the algorithmic model. In doing so, she has the negative task responsibility to not inflict harm by falsely accusing innocent citizen-clients or disproportionate burden on those found to have provided incorrect information. She also has a legally defined negative task responsibility not to base her judgment on protected or otherwise discriminatory indicators, such as ethnicity. The room for discretion and potential consequences in deviating from the algorithms risk indication are unknown. However, as she was epistemically dependent on the algorithmic model, there is reason to believe that this discriminatory risk assessment impacted her judgment of the application. Building on Van den Hoven's [23] argumentation, the civil servant also has the self-monitoring responsibility to check the algorithmic model, which automatically assigns applications to her. She has an obligation to raise concerns about the algorithmic model. Table 7 (Appendix) summarizes the civil servant's responsibilities.

We have already outlined how the design of the algorithmic model undermined the civil servant's conditions to fulfill these responsibilities (see Table 6, in the Appendix). However, the civil servant's autonomy in the decision-making process was further undermined by the institutional context in which the algorithmic decision-making process was embedded. As just a small cog in the larger 'benefits machine', civil servants were to follow a "zero-tolerance approach" or "all-or-nothing" approach [34]. Citizen-clients were forced to pay back benefits in full for the past up to five years for minor errors, such as incomplete information or missing signatures. Those citizen-clients who had to repay more than 3.000 Euro were automatically labeled with "deliberate intent or gross negligence" without further verification. These citizen-clients were no more eligible for debt payment plans or payment in installments. While this harm is primarily based on the harsh administrative sanctioning and enforcement policies, which are not the focus of this research, the algorithmic model must be judged as part of a socio-technical decision-making system. While the intention for implementing the algorithmic model was efficient and predictive decision-making at large scale, the algorithmic model has failed to indeed make correct decisions or to contribute to better decision-making procedures. The Dutch government concluded that between 2012 and 2019, 25.000-35.000 individuals were labeled as "deliberate intent or gross negligence" and, thus, had to repay the total amount of the benefits received. However, it was found that 94% percent of those judgments were incorrect.

Civil servants at the Allowances department fulfilled their self-monitoring responsibilities and raised concern about the workings and impact of this decision-making procedure with their management. A manager at the Allowances department noted that he encountered 'resistance' by the civil servants over the harmful impact this procedure has and their perception of being unable to 'do something for those impacted'. The legal advice of the Allowances department concluded that the tax authorities were merely implementing existing laws [34]. The civil servants' concerns were left unresolved.

4.3 The citizen-clients task responsibility in the application of the algorithmic model

This "zero-tolerance approach" placed much responsibility on the shoulders of the individual applicant [31]. Table 8 (Appendix) summarizes these responsibilities. However, the responsibility conditions necessary to take that responsibility were not intentionally designed in, and, in some cases, they were actively undermined, see Table 9

The citizen-client has the task responsibility to submit timely and correct (to the best of his or her knowledge) applications for childcare benefits. However, citizen-clients who were not sufficiently familiar with the Dutch language were not empowered by the administration to understand the applications. After information was published in different languages on the website, the error rate dropped, which subsequently led to dropping non-Dutch nationality as fraud indication [7]. There is an indication that many faults were not with fraudulent intention, but errors made in good faith. The applicant also has the crucial self-monitoring responsibility to satisfy herself that her application for childcare benefits is being processed accordingly. She is responsible for requesting information, challenging, and appealing decisions perceived to be epistemically or morally wrong. However, crucial responsibility conditions are curtailed. Citizen clients were not informed over the processing of their information by an algorithm. The citizen-clients were not provided with sufficient information about why their applications were dismissed or labeled fraudulent. There was no system of redress in place to provide citizen-clients with sufficient channels for challenging the decision. The administrative system to request information or appeal a decision is complex, centralized and citizen-clients received little to no information or support. Especially, those "economically and culturally distant from administrative rules and procedures" [31] found the complaints procedures challenging. As a citizen, one may be attributed a supervisory responsibility to voice wrongdoings and harm through government decision-making. However, many of those unjustly and untruly targeted do not have the socio-economic capacities to organize. The citizen-clients were not further empowered by other government organizations to do so.

Neither the responsibility conditions of the civil servants as end-users of the algorithmic model nor the responsibility conditions of the citizen-clients were deliberately designed for but undermined or compromised in the current socio-technical decision-making system. In retro perspective, we can only argue normatively that the tax authorities, ministries, and policymakers had an obligation to see to it that these responsibility conditions were brought about. The lack of explicitly transparently and continuously designing for responsibility is reflected in the understanding of the algorithm as the core technology, rather than it being part of a broader socio-technical system.

5 DISCUSSION & CONCLUSION

The application of task- and meta-task responsibility with the fine-grained responsibility vocabulary discussed above, and the attention to the responsibility conditions provide a more comprehensive conceptual lens to understand how the absence of explicit designing for responsibility in complex digital socio-technical systems enables harm and injustices in the application of AI in government decision-making. The analysis has enabled us to explore problems and provide answers to the research questions we posed up-front:

Which responsibilities should be allocated when using AI in governmental decision-making?

The five different task-related responsibilities (task responsibility, negative task responsibility, self-monitoring responsibility, supervisory responsibility, and meta-task responsibility) provide an analytical lens to identifying actor's various obligations and duties of a range of actors regarding the design, development, and application of artificial governance in public sector decision making. The task- and meta-task responsibility enable the design for responsibility within and across all levels of granularity. An actor's task responsibility may be developing a single system component or the task responsibility for adopting a legal framework. This approach enables the design for responsibility of socio-technical systems rather than isolated technological artifacts.

Which conditions need to be satisfied for the adequate attribution of these responsibilities?

The analysis has illustrated the relevance of establishing the necessary responsibility conditions – moral agency, freedom, control, and knowledge over one's contribution to the outcome – to be able to take and to be held morally responsible. For the developers, end-users, and citizen-clients, the adequate responsibility conditions to be held sufficiently responsible for (part of) their actions and inactions were not established or undermined. Consequentially, it would be unfair and ineffective to assign moral responsibility to these actors based on their inability to fulfill their task responsibilities and their involuntary contribution to a harmful outcome. If there is consensus that individuals a, b, c should be held morally responsible for X, Y, Z, we should design their artificial epistemic niche so that the conditions for responsibility are satisfied and holding them responsible is fair. Our proposed conceptualization enables a more thorough assessment and the identification of distinct conditions for specific task responsibilities. It must be acknowledged that though these conditions may guide design practices, they remain rather broad. In practice these conditions have to be contextualized and monitored in the continuous (re-)design of AI systems.

How do we design to ensure these conditions?

The retro perspective analysis highlights the importance of designing for responsibility upfront. It was epistemically impossible to attribute these lacking responsibility conditions satisfactorily to another actor's meta-task responsibility in retro-perspective. Despite having argued for an individualist notion of responsibility, we had to resort to attributing task responsibilities vicariously to government organizations and administrative units. This was since there was no design for individual moral responsibility in the first place. Eventually, we cannot satisfactorily distribute the individual moral responsibility for the harm caused by the application of the algorithmic model in the decision-making processes.

These findings further emphasize the importance of designing for responsibility ex-ante. Doing so will require attention not only to the artifact in isolation but as socio-technical systems embedded in a broader political and institutional context. The concept of meta-task responsibility as a corollary of task responsibility highlights the importance of designing for oneself and others' responsibility conditions. This conceptualization enables us to adopt an interconnected and holistic understanding of these responsibilities and, thereby, include AI systems' micro, meso, and macro context. This case shows that not only those immediately involved in the design and application of the algorithmic model bear task and meta-task responsibility. Instead, responsibilities, roles, and tasks are nested in further administrative and political institutions, which shape the preconditions of these actors as moral agents. These institutions, such as here the ministries, parliament, Cabinet, and the prime minister, bear task responsibilities themselves and meta-task responsibilities in ensuring the responsibility conditions of those downstream. This case exemplifies this in the political pressure exerted downwards to prioritize values such as efficiency, and the focus on combating fraud impacted the moral agency of those who implemented these policies.

These insights need to be considered in light of some limitations that reveal directions for further research. First, the case study analysis relies solely on reports provided by the government and non-governmental organizations. This approach made it difficult to distinguish responsibilities between and within different administrative units or the attribution to individuals. These documents are mainly focused on responsibility as accountability, liability, or blame. However, this paper does not intend to serve as an annex to a government inquiry. Instead, it seeks to explore whether and how designing for task responsibility can help understand past wrongdoings and prevent future harms by foregrounding the need for a fine-grained responsibility vocabulary that matches the complexity of the systems it is applied to. In our opinion, this is what is needed for a responsible design of algorithmic decision-making in government. Second, the conceptualization was not indented and, thus, fails to attribute complete backward-looking responsibility for the harm inflicted by the algorithmic model. This analysis shows both the relevance and difficulty of analyzing the socio-technical system rather than individual human or technical components. However, the task and meta-task responsibility conceptualization take the actors' tasks and necessary conditions as a starting point to design for responsibility and prevent future harm. Future research should engage practitioners in the co-creative design for the responsibility of algorithmic decision-making in government. Identifying and including all relevant actors, those operating the systems, and citizen-clients will mainly contribute to this design challenge. Future research should focus on integrating private technology companies as designers and users of AI in the public sector. In this paper we focus on the adequate attribution of responsibility-as-obligation. However, one may argue that the adequate attribution of responsibility is not necessarily effective in stimulating desirable behavior and, thus, preventing harm. Future research should focus on the relationship between adequate or fair and effective attributions of responsibility. Theoretical and conceptual contributions from public accountability studies may contribute to closing this gap.

The use of algorithms in socio-technical decision-making systems requires us to rethink the allocation of responsibilities and the conditions needed to take these responsibilities. The deliberate and inclusive design for responsibility can prevent governments from "stumbling, zombie-like, into a digital welfare dystopia" [6]. Instead, AI can contribute to more responsive, just, and effective social welfare systems if designed for responsibility. In AI and socio-technical systems that will be increasingly introduced in public administration, we cannot hold people responsible if they have not been made responsible, and our design mirrors this. They cannot be held responsible if the epistemic conditions do not support (or even undermine) them in taking moral responsibility. We should not expect to answer questions about who was responsible in AI and Data Government applications in a morally satisfactory way if the context and the applications were not adequately designed.

REFERENCES

[1] European Commission. 2022. Excellence and trust in AI. European Commission. Retrieved January 27, 2022, from https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/excellence-trust-artificial-intelligence_en.

[2] Anneke Zuiderwijk, Yu-Che Chen, and Fadi Salem. 2021. Implications of the use of AI in public governance: A systematic literature review and a research agenda. Government Information Quarterly, 101577.

[3] Bernd Wirtz, Jan Weyerer and Carolin Geyer. 2019. AI and the public sector—applications and challenges. International Journal of Public Administration, 42(7), 596-615.

[4] Bernd Wirtz, Jan Weyerer and Benjamin Sturm. 2020. The dark sides of AI: An integrated AI governance framework for public administration. International Journal of Public Administration, 43(9), 818-829.

[5] Tara Qian Sun and Rony Medaglia. 2019. Mapping the challenges of AI in the public sector: Evidence from public healthcare. Government Information Quarterly, 36(2), 368-383.

[6] Philip Alston. 2019. Report of the Special Rapporteur on extreme poverty and human rights (A/74/493). United Nations General Assembly. https://documents-dds-ny.un.org/doc/UNDOC/GEN/N19/312/13/PDF/N1931213.pdf?OpenElement

[7] Autoriteit Persoonsgegevens. 2020. Belastingdienst/Toeslagen: De Verwerking van de Nationaliteit van Aanvragers van Kinderopvangtoeslag. 1–60.

[8] Tweede Kamer der Staten-Generaal. 2020. 35 510 Parlementaire ondervraging Kinderopvangtoeslag: Brief Van De Parlementaire Ondervragingscommissie (35510 nr. 2). Tweede Kamer der Staten-Generaal. https://zoek.officielebekendmakingen.nl/kst-35510-2.html.

[9] Matthew Miles and Michael Huberman. 1994. Qualitative data analysis: An expanded sourcebook. Sage.

[10] Changlin Wang, Thompson SH Teo, and Marijn Janssen. 2021. Public and private value creation using AI: An empirical study of AI voice robot users in Chinese public sector, International Journal of Information Management, Volume 61, 2021.

[11] Jack Stilgoe. 2018. Machine learning, social learning and the governance of self-driving cars. Social studies of science, 48(1), 25-56.

[12] Robert Kitchin. 2017. Thinking critically about and researching algorithms. Information, communication & society, 20(1), 14-29.

[13] Andrew Selbst, Danah Boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, and Janet Vertesi. 2019. Fairness and abstraction in sociotechnical systems. In Proceedings of the conference on fairness, accountability, and transparency (pp. 59-68).

[14] Nicole Vincent. 2011. A structured taxonomy of responsibility concepts. In Moral responsibility (pp. 15-35). Springer, Dordrecht.

[15] Herbert Lionel Adolphus Hart. 1968. Punishment and Responsibility: Essays in the Philosophy of Law. Oxford: Clarendon Press.

[16] Ibo van de Poel. 2015. The problem of many hands. In Moral responsibility and the problem of many hands (pp. 62-104). Routledge.

[17] Ibo van de Poel and Martin Sand. 2018. Varieties of responsibility: two problems of responsible innovation. Synthese, 1-19.

[18] John Martin Fischer and Mark Ravizza (Eds.). 1993. Perspectives on moral responsibility. Cornell University Press.

[19] R. Jay Wallace. 1994. Responsibility and the moral sentiments. Harvard University Press.

[20] Ibo van de Poel. 2011. The relation between forward-looking and backward-looking responsibility. In Moral responsibility (pp. 37-52). Springer, Dordrecht.

[21] Robert E. Goodin. 1987. Apportioning responsibilities. Law and Philosophy, 6(2), 167-185.

[22] Robert E. Goodin. 1995. Utilitarianism as a public philosophy. Cambridge University Press.

[23] Jeroen van den Hoven. 1998. Moral responsibility, public office and information technology. Public administration in an information age: a handbook, 97-112.

[24] Neelke Doorn. 2012. Responsibility ascriptions in technology development and engineering: Three perspectives. Science and Engineering Ethics, 18(1), 69-90.

[25] Alan Rubel, Clinton Castro, and Adam Pham. 2019. Agency laundering and information technologies. Ethical Theory and Moral Practice, 22(4), 1017-1041.

[26] Jeroen van den Hoven, Pieter E. Vermaas, and Ibo Van de Poel. 2015. Design for the value of responsibility. Handbook of ethics, values, and technological design. Springer, Dordrecht, 473-490.

[27] Dennis F. Thompson. 1980. Moral responsibility of public officials: The problem of many hands. American Political Science Review, 74(4), 905-916.

[28] Mark Bovens. 1998. The quest for responsibility: Accountability and citizenship in complex organisations. Cambridge university press.

[29] Dennis F. Thompson. 2014. Responsibility for failures of government: The problem of many hands. The American Review of Public Administration, 44(3), 259-273.

[30] European Commission For Democracy Through Law (Venice Commission). (2021, October). The Netherlands Opinion On The Legal Protection Of Citizens (No. 1031/2021). Council of Europe. https://www.venice.coe.int/webforms/documents/default.aspx?pdffile=CDL-AD(2021)031-e

[31] Netherlands House of Representatives, Verslag - Parlementaire Ondervragingscommissie Kinderopvangtoeslag: Ongekend Onrecht, 17 December 2020, p. 1, tweedekamer.nl/sites/default/files/atoms/files/20201217_eindverslag_parlementaire_ondervragingscommissie_ kinderopvangtoeslag.pdf.

[32] Alexandra van Huffelen. 2020. Kamerstuk II 2019/20, 31 066, nr. 683: https://zoek.officielebekendmakingen.nl/kst-31066-683.html

[33] Amnesty International. (2021, October). Xenophobic Machines - Discrimination through unregulated use of algorithms in the Dutch childcare benefits scandal. https://www.amnesty.nl/actueel/xenophobic-machines-discrimination-through-unregulated-use-of-algorithms-in-the-dutch-childcare-benefits-scandal.

[34] Advisory committee on implementation of benefits, Omzien in verwondering 2: Eindadvies Adviescommissie Uitvoering Toeslagen, 12 March 2020, p. 48, rijksoverheid.nl/documenten/kamerstukken/2020/03/12/omzien-in-verwondering-eindadvies-adviescommissie- uitvoering-toeslagen.

[35] Alexandra van Huffelen. 2021. Kamerbrief over openbaarmaking risicoclassificatiemodel Toeslagen (Kamerstuk 31066, nr. 923).

[36] Luciano Floridi and Jeff Sanders. 2004. On the morality of artificial agents. Minds and machines, 14(3), 349-379.

[37] Luciano Siebert, Maria Luce Lupetti, Evgeni Aizenberg, Niek Beckers, Arkady Zgonnikov, Herman Veluwenkamp, David Abbink *et al.* 2021. Meaningful human control over AI systems: beyond talking the talk. arXiv preprint arXiv:2112.01298.

APPENDIX

Table 5: Task responsibilities – What are the task responsibilities the developers of the algorithmic model are attributed?

Task responsibilities – What are the task responsibilities the designers of the algorithmic model are attributed?	
Task responsibility	The Allowances department was responsible for the operational task of granting, paying, and recovering childcare allowances. With the political priority being the efficient prevention and combat of fraud in the childcare benefits system, they were to check the applications before payments were made, particularly of those without prior residence in the Netherlands. To do so more time and cost efficiently, they were tasked with developing an ICT system that was able to screen the applicants for fraud automatically and on a large scale.
Negative task responsibility	The developers had the negative task responsibility to do so without inflicting harm on innocent, non-fraudulent applicants, as well as reducing the costs of the ICT system.
Self-monitoring responsibility	The developers had the self-monitoring responsibility to monitor and, if necessary, re-design the self-learning algorithmic decision-making system. This includes the responsibility to design the algorithmic model so that it allows for continued oversight and an understanding of its internal workings.
Supervisory responsibility	The developers had the supervisory responsibility to see to it that the end-users of the model use the risk score as an indication for manual assessment yet are still able to assess the application independently.
Meta-task responsibility	The developers had the obligation to see to it that the end-user can use the algorithmic model responsibly. By designing for the end-user's responsibility conditions, she is empowered to take responsibility for the outcomes of her application of the algorithmic model (see Table 6).

Table 6: The developers meta-task responsibility to design for the end-user's responsibility conditions

Designing for operator's responsibility conditions – When is adequate to attribute responsibility to the operator?		
Designing for moral agency & intentionality	Designing for freedom & control	Designing for knowledge
The operator must be able to work with the system as a morally autonomous person and understand her model-independent actions.	The operator must be (and perceive to be) free in the application of the algorithmic model and possess a professional room for discretion to act accordingly.	The end-user should understand why and how the algorithmic model arrives at a certain risk-indication.
The developers neglected their meta-task responsibilities to design for the operator's responsibility conditions. The design of this algorithmic model makes the operator epistemically dependent upon the model's output. The design of the broader socio-technical decision-making processes further designs out their room for discretion.		

Table 7: Task responsibilities – What are the task responsibilities attributed to the civil servant as an end-user of the algorithmic model?

Task responsibilities – What are the task responsibilities attributed to the civil servant as operator of the algorithmic model?	
Task responsibility	The civil servant has the task-responsibility to assess those applications that are flagged as high-risk by the algorithmic model.
Negative task responsibility	The civil servant has the obligation to independently assess each flagged application. She must do so without inflicting harm or false accusations on innocent applicants, this includes no basing one's decision on protected or otherwise discriminatory indicators, such as ethnicity.
Self-monitoring responsibility	The civil servant has the obligation to check the algorithmic model which automatically assigns cases to her. She also has to see to it that no disproportioned burdens are placed on incorrect or potentially fraudulent applicants through her manual assessment. If she does perceive such harm, she has the obligation to act and raise her concern.
Supervisory responsibility	The civil servant has the supervisory responsibility to see to it that the applicants provide correct and refrain from providing incorrect or fraudulent information.
Meta-task responsibility	As the end-user who is not involved in the design of the algorithmic model or broader decision-making system, the end-user is neither attributed meta-task responsibilities, nor has the power to impact the relevant conditions.

Table 8: Task responsibilities – What are the task responsibilities the applicant (citizen-client) can be attributed?

Task responsibilities – What are the task responsibilities attributed to citizen-client?	
Task responsibility	The citizen-client has the task responsibility to submit a timely and correct (to the best of his or her knowledge) applications for childcare benefits.
Negative task responsibility	In applying for childcare benefits the applicant has an obligation to ensure that a correct application avoids large reclaims.
Self-monitoring responsibility	The citizen-client has the self-monitoring responsibility to satisfy herself that her application for childcare benefits is being processes accordingly. She has the responsibility to request information and challenge wrong or unjust decisions.
Supervisory responsibility	A citizen-client is also a free citizen in a liberal-democratic system and, thus, has the supervisory responsibility to voice unjust treatment and immoral behavior by those elected or appointed to power.
Meta-task responsibility	The citizen-client has the meta-task responsibility to ensure her own responsibility conditions, as it is within her own capabilities.

Table 9: Responsibility conditions of the citizen-clients were not met, making it impossible to attribute responsibility fairly and effectively.

Responsibility conditions of the citizen-client were not met make it impossible to adequately attribute responsibility		
Designing for moral agency & intentionality	Designing for freedom & control	Designing for knowledge
The citizen-client must be able to act intentionally.	The citizen-client must be able to raise her concerns in case she feels unjustly treated.	The citizen-client must understand his responsibilities to provide correct information and must understand the potential consequences or incorrect information.