AI & SOCIETY
https://doi.org/10.1007/s00146-024-01920-4

MAIN PAPER

Citizens' data afterlives: Practices of dataset inclusion in machine learning for public welfare

Helene Friis Ratner[1,2] · **Nanna Bonde Thylstrup**[2]

Received: 2 October 2023 / Accepted: 5 March 2024

Abstract
Public sector adoption of AI techniques in welfare systems recasts historic national data as resource for machine learning. In this paper, we examine how the use of register data for development of predictive models produces new 'afterlives' for citizen data. First, we document a Danish research project's practical efforts to develop an algorithmic decision-support model for social workers to classify children's risk of maltreatment. Second, we outline the tensions emerging from project members' negotiations about which datasets to include. Third, we identify three types of afterlives for citizen data in machine learning projects: (1) data afterlives for training and testing the algorithm, acting as 'ground truth' for inferring futures, (2) data afterlives for validating the algorithmic model, acting as markers of robustness, and (3) data afterlives for improving the model's fairness, valuated for reasons of data ethics. We conclude by discussing how, on one hand, these afterlives engender new ethical relations between state and citizens; and how they, on the other hand, also articulate an alternative view on the value of datasets, posing interesting contrasts between machine learning projects developed within the context of the Danish welfare state and mainstream corporate AI discourses of the bigger, the better.

Keywords Machine learning · Welfare state · Data afterlives · Dataset negotiations

1 Introduction

Public sector bodies increasingly adopt machine learning techniques as part of governing schemes across health, social services, policing, and employment services (Dencik et al. 2019; Hoeyer 2019; Jørgensen 2023). The machine learning projects emerging out of these initiatives reconfigure historical national data into a dynamic resource for future-oriented algorithmic profiling of citizens' behavior, rights, and needs. With machine learning, correlations found in historical data are generalized into rules, often in the form of algorithmic prediction of citizens as part of service provision in public administration. In this sense, citizen data from cases now closed, decisions and allocations long forgotten and expired in municipal casework, acquire a new existence in emerging profiling and predictive models.

Drawing upon the theoretical framework of *afterlives* as it has been developed in relation to data, digital infrastructures, and archival regimes (Agostinho 2019; Ebeling 2022; MacKinnon 2022; Sutherland 2023), we aim to elucidate the complex trajectories through which data continue to exert influence and significance after their initial collection and utilization. This analytical approach allows us to investigate the dynamic interplay between past and present data practices, and explore future implications of these within the realm of family welfare services.

Empirically, we document a Danish research project's practical efforts to develop an algorithmic model for classifying children's risk of maltreatment through prediction. Envisioned as a decision-support tool, the research project investigated whether such a model can improve caseworkers' risk assessments through the algorithmic attribution of risk scores to children. Examining negotiations about which citizen data to include when developing the model, i.e., negotiations about the extent and form of data afterlives, we identify how different concerns shape the researchers' inclusion

✉ Helene Friis Ratner
helr@edu.au.dk

Nanna Bonde Thylstrup
nannab@hum.ku.dk

[1] Danish School of Education (DPU), Aarhus University, 2400 Copenhagen N, Denmark, Tuborgvej 164

[2] Department of Arts and Cultural Studies, University of Copenhagen, Karen Blixensvej 1, 2300 Copenhagen, Denmark

Published online: 10 April 2024

or exclusion of different data variables. This examination extends into how data scientists navigate tensions across domain knowledge, machine learning expertise, legal assessments, and medialized public critique in this development.

Arguably, the repurposing of archival data is not novel, especially not in a Denmark, internationally renowned for comprehensive registers with access to detailed information on the population's health, demographics, and socioeconomic factors (Erlangsen and Fedyszyn 2015). As such, Danish archives, containing wide-ranging records of past state–citizen interactions, have long been used for epidemiology and other types of register research, including for the development of evidence-based policy. Yet, we propose that data afterlives in a context of machine learning, potentially used for profiling at the level of the individual rather than the population, implies a new type of data afterlife. As we will demonstrate in our analysis, inclusion of archival data in machine learning projects gives citizen data a new digital existence. In this article, we identify three types of data afterlives and highlight the complexities of dataset inclusion in machine learning in a context of Scandinavian welfare institutions.

The structure of this paper unfolds as follows: initially, we present the empirical foundation underpinning our research. Subsequently, we offer a review of literature pertaining to input data in machine learning models, with a particular emphasis on critical dataset studies and Science and Technology Studies (STS). Following this, we present the concept of data afterlives to elucidate the ethical and political dimensions inherent in the repurposing of historical data for predictive analytics. With this concept, we elicit three distinct iterations of citizen data afterlives within machine learning models, leveraging insights gleaned from both empirical observations and interview data. Lastly, we provide a conclusive section delineating potential avenues for future research inquiries into machine learning initiatives within the public sector.

2 Empirical resources

This paper reports a long-term (2021–2023) case study of RISK, an interdisciplinary Danish research project investigating an algorithmic approach to family welfare services. The project is privately funded and involves collaboration between a university with expertise in statistics and machine learning and a university college focusing on social work. In this sense, it builds on a well-established model for organizing the heterogeneity of expertise, 'the logic of domains,' where machine learning is cast as the universal, domain-independent technique that can serve domain-specific sciences such as social work, through the introduction of new exploratory techniques and tools (Ribes et al. 2019).

The purpose of the RISK project was to determine if it was feasible to create a predictive algorithm with sufficient accuracy to be used as a decision-support tool in case worker assessments of children referred to the social services. In Denmark, municipalities by law are obligated to assess referrals within 24 hours with regards to the acuteness and severity of the child's situation, an assessment that has been documented to be fraught with uncertainty and a lack of standardization (Villumsen and Søbjerg 2020). Part of the research project was also to assess the impact on case workers and families whose children had been evaluated by the algorithm. In interviews, the project partners emphasized the importance of having interdisciplinary research on the effects of algorithmic predictive models before making decisions about their implementation.

While algorithmic decision-support models in child protection services are beginning to gain ground internationally, they have not yet been implemented in Denmark due to a lack of legality and high degree of controversiality (Eubanks 2018; Leslie et al. 2020; Redden et al. 2020). The controversiality stems from a previous attempt to develop a different predictive algorithm in child protection services and regarded the potential discriminatory nature of algorithmic profiling (e.g., if the model generates biased outputs) as well as the risk of undue surveillance (Kristensen 2022). At the same time, the very field of child protection work is fraught with competing values and is politically contested, not least due to Danish Prime Minister Mette Frederiksen's call for municipalities to make more use of forced adoptions and out-of-home placements in the name of child protection (Frederiksen 2020). These events have produced a Danish context where machine learning experiments in child protection services have much public scrutiny (Ratner and Schrøder 2023). For RISK, public contestations led to uncertainty about the legality of their algorithmic model and, as a consequence, the researchers have abandoned testing it on real-life cases. Currently, its viability as a future technology thus looks unlikely. For our purpose, however, these contestations are important insofar as they have led to new decisions about how to include data variables in the development of the model. In that sense, they reflect how public engagement in science has impact on decisions about data set inclusion, and ultimately, citizens' data afterlives.

The algorithmic model of RISK is trained using pseudonymized national registry data, which can only be accessed by researchers who have obtained ethical approval. The central data set comes from Statistics Denmark's archive for referrals. The primary purpose of Statistics Denmark is to collect data and compile and publish statistics on the Danish society. These data include a wide range of data types such as population data, data from municipalities, private sector companies, cultural institutions, and the banking sector. The archive for referrals

was established in 2015 as part of the 'assault package' (2013), a legislative change to the Social Service Law, which should enhance protection of children by strengthening municipalities' handling of referrals. It required all Danish municipalities to submit annual data about referrals of children under the age of 18, including information about the age and gender of the child, the date of referral, the referrer's relationship to the child (e.g., teacher), the reason for referral, and municipality. The assault package was introduced in the aftermath of medialized scandals about severe sexual abuse of children where municipalities had failed to act on referrals. The then social and interior minister Karen Elleman described the new archive as an:

> overview over the referrals regarding children at risk sent to municipalities and how municipalities follow up (…) I hope that each municipality will use this new tool to examine their own practice. If, for example, one receives significantly fewer referrals than they do in the neighbouring municipality, does this generate an occasion to examine why this is the case? (Elleman 2015).

The hope at the time, thus, was to render visible through comparison if municipalities had suspiciously few referrals or had failed to act on these, made visible and comprehensible through numerical comparison. As data afterlives, referrals contain traces of children's and families lived experiences and encounters with welfare institutions within health, childcare, and education—and the concerns that have materialized with welfare professionals stewarding these institutions. These concerns become datafied in the form of referrals, assessed by case workers in municipal family welfare services, eventually taking up residence in the national archive of referrals.

RISK's main dataset comes from the national archive of referrals. The researchers trained their algorithmic model on all nationally registered referrals concerning children from April 2016 to December 2017, which amount to 173,044 referrals about 90,644 children. From Statistics Denmark, they also include register data from child protection services about the children and their families for 2 years prior to and 1 year after the 2016–2017 period. This allows them to explore relationships between referrals and decisions made or not made by the family welfare services. The algorithm is trained using this extensive scope of data to predict the probability that a child with a referral will face maltreatment and adverse development in the coming year. As a proxy for child maltreatment, the model utilizes the subsequent placement of the child in out-of-home care within the year following the referral as outcome measure (Report on the Statistical Model, 2021). In terms of afterlives, we are thus looking into the repurposing of referrals and casework, once written to help the child and family in question, and now being used to develop a predictive model.

3 Social studies of machine learning: from outputs to inputs

In recent decades, there has been a notable increase in interdisciplinary research examining the risks and opportunities associated with public sector initiatives in machine learning concerning citizens and vulnerable populations (Plesner and Justesen 2022; Ranchordas 2021; Ratner and Elmholdt 2023). Much of this scholarly inquiry has concentrated on the outcomes of data assemblages, such as enhanced efficiency, governmental paradigms, and algorithmic discrimination. However, more recently, a growing body of work within *critical dataset studies* (Thylstrup 2022) has also begun to explore the *input* aspect of machine learning including training, benchmarking and test data.

Critical research on data sets is particularly influenced by two main areas of study: AI accountability and ethics, and social studies focused on machine learning technologies. In the realm of AI ethics, scholars have challenged the prevalent notion within computational sciences that data sets are merely tools for digital knowledge production. Instead, they argue that data sets should be viewed as archives of socio-cultural data (Jo and Gebru 2020). This perspective shift has sparked a surge of research examining the political implications of data set origins, standards for data collection and usage, and the underlying values shaping data set development (Denton et al. 2020; Paullada et al. 2020; Raji et al 2021; Scheuerman et al 2021). More recently, there has been a growing emphasis on data set auditing as a key area of inquiry within this field (Raji & Buolamwini 2022). Collectively, these studies aim to establish new fairness and justice standards in AI contexts, along with novel methodologies for holding AI systems accountable through documentation and auditing practices.

A second strand of research on data sets emerges out of the intersection of STS and neighboring fields such as media, philosophy and history of science and communication. This body of work is concerned with the 'data paradoxes' (Hoeyer 2023), data-driven practices and machine learning assemblages, including the "science frictions" (Edwards et al 2011) and epistemic cultures (Leonelli and Tempini 2020) that shape machine learning environments and practices. Important work within this body focuses on data set practices and epistemologies (Plantin 2019; Ratner and Ruppert 2019; Walford 2013); the social relations between data sets and models in machine learning (Ribes et al. 2019); the implications of increased data set interoperability and sharing (Ribes 2017; Slota et al. 2020; Thylstrup et al. 2022), and the assumptions of domain independence

that shape much data scientific practice (Hansen and Borch 2022; Ribes et al. 2019; Hartley Møller and Thylstrup 2024). Especially relevant to this study is the interest in exploring how data sets are assigned different roles within the construction workflow of algorithms in machine learning and the uncertainties that shape these decisions (Jaton 2021) as well as the genealogies that training data sets, testing data sets and—sometimes—also benchmarking data sets emerge out of (Hanna et al. 2020). The synthesis of these two research streams highlights the importance of not only examining the effects of algorithms but also questioning the input processes when assessing ethical implications of machine learning models. This encompasses considerations of extraction, selection, and interpretation methodologies, as these factors are intricately linked with larger power dynamics and knowledge frameworks.

Drawing on these perspectives, the upcoming section introduces the concept of data afterlives to elucidate how context-specific social interactions, influencing decisions about data input, intersect with broader ethical considerations concerning the relationship between the state and its citizens. Our aim is to demonstrate that dataset practices entail more than just technical proficiency; it encompasses issues of ethical considerations, "science frictions" between different fields of expertise, and, particularly in the case of algorithms intended for potential deployment in the public sector, the emergence of new dynamics in citizen–state relationships.

4 Conceptualizing input data as data afterlives

Scholars in the social sciences and humanities have long been interested in the dynamic nature and vitality of social and non-social data (Nadim 2016; Kaufmann & Leese 2021; Medina Perea et al. 2020; Winthereik 2023). Moreover, scholars within the field of digital studies have also paid an interest in the ways in which digital technologies and networks are inflected with human remains (Jucan et al 2019; Thylstrup et al. 2022; Sutherland 2023) as well as the ethico-political issues that accompany archaeological excavation of these traces (Agostinho 2019; Odumosu 2020; Mackinnon 2022). The latter strand has also been instrumental in fostering attention to the ways in which archival regimes enfolds pasts and futures within the present in ways that give rise to new ethical considerations and contestations (Schneider 2011; Keenan 2018).

This paper draws on these perspectives, mobilizing the concept of data afterlives to elucidate how the social practices shaping data set inclusion within public sector machine learning initiatives give rise to new datafied relations between citizens and the state. We believe that the RISK project offers one very concrete example of these transtemporal data relations as it transforms the epistemologies of past records to signals of future citizen risk.

As Louise Amoore (2020: 11) points out, algorithmic systems 'modify themselves in and through their recursive relations to input data' in such a way that '[l]ittle pieces of past patterns enter a training data set and teach the algorithm new things … on and on iteratively, recursively making future worlds'. While algorithmic systems such as RISK are not generative, they nevertheless illustrate this transtemporal relation, producing new entanglements between data, models and social systems that mutually constitute each other in recursive ways (Thylstrup et al. 2022). Understanding the dynamic nature of data and their afterlives is especially relevant within the context of family welfare because the very category of child maltreatment itself has changed with history. As Ian Hacking points out in "The making and molding of child abuse" (1991) the very idea of child abuse and neglect has been in "constant flux" over the past 50 years, and theories and methods of what counts as a signal have changed significantly over the years. It is also within this intricate transtemporal interplay between classification practices, predictive methods, scientific developments, political transformations and social categories (Hacking 1991) RISK unfolds, here constituting classification practices from 2016 to 2017 as a 'ground truth' for algorithmic predictions.

As the analyses in the following section show, data afterlives are shaped by conflicting understandings and practices of values and constraints (Heuts and Mol 2013; Lee and Helgesson 2020). By focusing on the moments of friction in the valuation of data, we show the uncertainties data scientists face when deciding which data sets to include and exclude. Such moments, we believe, offer crucial insights into the shaping of citizen data afterlives: which registers of value are involved in the decisions about which data to include or exclude in data set development? And which compromises are made when there are clashes between different registers (cf. Heuts and Mol 2013)? In asking these questions, we analytically attend to constraints and demands coming from both within the epistemic conventions associated with machine learning as well as those constraints which emerge from other sources.

5 Analytical section: examining data afterlives in machine learning systems

5.1 Citizen data afterlives as ground truth

A central element in the transformation of citizens' data afterlives through machine learning is *ground truthing*, i.e., compiling a database with a clear labeling of outputs (measurements of children's serious maltreatment), which

the desired algorithm should be able to predict (cf. Jaton 2017). In fact, it was exactly the pathbreaking and applied nature of quantitative method within the domain of risk assessments that appealed the junior data scientist (JDS) on the team: "I could see that this is a direct application of the work I will be doing' (JDS, Oct, 2021). From the outset, the prospect of doing research that could be implemented in practice and potentially help children was a key motivation for joining the project. In characterizing the merits of a machine learning approach to risk assessment, compared to existing assessments practices, JDS described the quantitative approach as a broadening of analytic possibilities within the risk assessment apparatus because it involves an expansion of citizen data points beyond the data made available to the individual case worker. JDS articulated this breadth of data points in terms of "completeness': 'We have the possibility of using referrals from the entire country, right? So maybe that is one of the big advantages of this project, right? That we have complete data' (JDS, Oct, 2021).

Completeness here is understood in sense of the referrals archive operated by Statistics Denmark. In a research article describing their datasets, they also note how their '[data] set contains the complete history of a child's past interactions (notifications, interventions, and placements) with all Danish CPS [child protection services].' [p. 8]. At the same time, there are no data for 'serious maltreatment' and they needed to find a proxy. As the senior data scientist (SDS) explained:

> So what we really wanted to do was to help the social worker identify the children they estimated to be in need of help from the government in the form of interventions or out of home placements. (…). And the problem is that "maltreatment' doesn't exist in the registers. I don't have a good target for that (…). Uhm, so our approach was, well let's try to see; is there something in the records that we don't expect is correlated, or that we expect is correlated with maltreatment. And there we look at out-of-home placements and we also look at future referrals and uhm interventions with the home etc. Uhm we found that (…) future out of home placements (…) had the best predictive capabilities. (SDS, May, 2021)

Much work went into compiling the ground-truth data set, in terms of deciding which data should serve as output target. To find a proxy for maltreatment, they did an additional statistical analysis on various possible targets and correlated them with other data points indicative of 'serious maltreatment', not to be used to train the algorithm (e.g., data on psychiatric illness, school absence). Thus, data regarding children's out-of-home placement acquired value as a proxy from its correlation with other targets deemed to be associated with maltreatment.

Once the ground-truth data set has been established—and the output target decided—the research team tested different machine learning algorithms to develop a predictive algorithm. In this process, they divided the data set into two:

> And then we have developed the model on 70% of these data. And then we have tried to see, well, how does the performance apply to the last 30%? Which means that the final 30% plays the role as the new cases, which the algorithms didn't see when it was being developed. So, what we do is that we develop the model on 70% of our data. And then we evaluate it on the final 30%. Because in this way we can be sure that we didn't overfit our data, but that we developed a model that can catch some general tendencies. (JDS, Oct, 2021)

Developing the predictive algorithm, i.e., generating its capacity to infer futures, in other words depends on dividing the ground-truth database in two, one for training the algorithm and one evaluating its performance on the rest of the ground-truth data. The first kind of data set is called *training data* and data scientists use this data to 'teach' computers to recognize a pattern. This data set, thus, also forms the foundation of what a model will subsequently know of the world. The second set of data—testing data—is used to evaluate how well a model predicts a particular feature, here out-of-home placements. As testing data, data acquire value in allowing data scientists to test how well a given model recognizes the feature it was trained to detect. This step is crucial because it allows the data scientist to gauge how well the model will work "in the wild'. Yet, the controlled test itself is also limited by the testing data.

As ground truth, citizens' data afterlives attain the status as a form of 'truth-telling' in machine learning practices. Describing ground truths as the 'referential repositories that work as material bases for algorithms' (2021: 24), Jaton stresses the negotiations and compromises involved in the ground-truthing practices that underlie the shaping and use of ML algorithms. In doing so, Jaton highlights a trivial but often forgotten feature of computer science research, whereby the consistency of an algorithm is established. Through his focus on the conventions and processes of ground truthing, Jaton, thus, not only shows the fragility and uncertainty of computer science in general but also the notion of ground truth in particular. Instead of seeing ground truth as a form of representation that directly transmits an unshakeable reality, it is rather a practice that involves several alternatives.

Such negotiations also took place here, not at least due to the public scrutiny the project was exposed to. As the junior data scientist explained, this led to the realization that

> I think that to begin with, I perhaps approached it a little like I was used to. Where you could say, okay, this is fun data (…) I think it kind of snuck in on me that I realized how different – that this is a much bigger responsibility than I ever had before. It has the public (...) what do you call it? [attention] (…) so of course you kind of need a justification for doing what we are doing. (JDS, Oct 2021)

Thus, the very transformation from archival data to ground-truth database was not simply a matter of including as many data sets as possible. Many different concerns were negotiated, including attention to the public's potentially critical reception of the project. While the latter aspect led to iterative reductions of data variables, from a data-ethical perspective and the understanding that all choices should be justified, there were also moments of data inclusion coming from the data scientist's 'intuition', i.e., cultural and popular understanding of what constitutes risk to children, and from his sparring with the research project's domain experts.

> Well, we don't need an unnecessary large amount of data (…) we have this sparring with especially [anonymized] municipality about the kinds of things they are looking at when they receive a referral. (…) And then we have tried to say, okay, is this something we can – is that something we can recreate in the registers? Uhm, and then there has of course been a lot of intuition about, well what kinds of things. (…) Even if I am hired as a calculator boy, I have perhaps also given some consideration to, well, what kinds of things that could have an implication for [maltreatment]. (…) I think, for instance, that something like parents' age at birth. At the birth of the child. I think – I don't know if I discussed it with so many people, but I just thought that it would be an obvious thing to take into account .(…). I maybe think that there is good evidence for the risk associated with being teenage parents. (JDS, Oct 2021)

From this perspective, ground-truthing practices emerge not only as a form of scientific practice that operates within its own epistemic paradigm but also is enveloped in a deeper societal construction of algorithmic ethics regarding data minimization as well as domain and non-domain ideas about what constitutes risk.

5.2 Citizens' data afterlives as variables for post-validation

In addition to generating ground truths for inferring futures, citizen data was also included to *validate* the model, giving rise to a different type of data afterlives. Decisions about whether citizens' datasets should be used for the ground-truthing database or for validating the model were partially shaped by externally imposed constraints and requirements that led them to reduce their ground-truth database. Rather than deciding not to use these data sets at all, the researchers assigned them a different role. Below, we examine the concerns that led the researchers to remove datasets from ground-truth database as well as their rationales for including them in post-validation practices.

From the onset, the legality of the algorithm was an important concern, not at least due to the original ambition of its possible implementation in municipal child protection services. Reflecting on the media controversy regarding the so-called Gladsaxe model, mentioned in the beginning of this article, the social work project leader (SWL), a psychologist, warned against developing a model that would be illegal from the onset. Instead, they highlighted the need for as small and delimited data set as possible.

> I mean, it is not cool if we develop a tool that over-informs a case, with the consequence that there is no legal basis for its implementation (…) and we are not allowed to merge data across administrative units [in municipalities] (…) Even if we are legally allowed to include all these data points simply for purposes of research. But then we end up with a tool which cannot be implemented because it's illegal (SWL, May 2021).

To secure legal compliance, they first, as we also saw in the previous analytical section, consulted domain expertise with their partner municipalities, looking into which data the social workers were already accessing, and developed the algorithmic model using these data variables, which they assumed to be legal. This meant that they did not include data from other administrative units. At the same time, they were curious to learn whether an inclusion of such illegal variables would enhance the model's precision, referring to this as a 'kitchen sink approach where you just throw everything into the model' (SDS, May 2021). This tension between the need for mathematical precision, using a traditional 'kitchen sink approach', and legal compliance, requiring them to limit data, was managed through iterations of post-validation analyses.

Thus, while datasets about citizens' health or education were not used to train the model, researchers decided to include them in the post-validation of the model. As these data would only be used as part of the research process, it did not impede on the eventual legality of the model as a decision-support tool. As the junior data scientist reflected. 'These variables have not been used to build the model as we are not allowed to use health data for our model. But we have subsequently (…) used data from the National Patient Register. We have used it for the post-validation of our model only' (JDS, October 2021). The decision-support tool, in other words, would not need to pull these data

to score children's risk of maltreatment. In this way, they could scope whether precision would increase with more data (only very little) without training the algorithmic model on illegal data sets.

In this practice of post-validation, they tested the model's predictive capacities against other adverse outcome targets than their original proxy for maltreatment (out-of-home placement), with the following rationale: 'if removal and subsequent placement generally occur only in extreme cases of maltreatment, higher risks of removal and placement are associated with greater risks of other adverse child outcomes, which are themselves indicative of maltreatment' (ANON draft article p. 7). These included children being charged with crime, fractures, mental illness, fraction of damaged teeth, and illegal school absence. Citizens' data afterlives, thus, also take a shape of testing the model's validity after it has been developed. As they justified in a scientific article, data sets such as fractures might even be a more valid proxy for children's maltreatment:

> Although child maltreatment only accounts for a small share of the total number of fractures, they are the second most common injury caused by physical abuse among children (bruises being the most common one) (38), and are frequent among young abused children with up to a third of them experiencing fractures (39, 40, 41). The main advantage of this outcome lies in the fact that the associated measurement error is plausibly less influenced by external sources such as CPS biases (and possibly households' reaction to actions taken by CPS)—unlike many other adverse events analyzed (ANON draft article p. 14).

In these ways, data sets, illegal to use in an ADM model, acquired a shadow existence for post-validation. They were deemed valuable because of they were seen to be less biased than data from child protection services, which involve both case worker interpretation and relational work on the side of social workers and families. Rather than data for ground truthing, prediction, these data afterlives figure as different form of truth used to confirm or reject the model's predictive power and hence to render it more robust.

5.3 Citizens' data afterlives as variables for (un)fair algorithmic profiling

Researchers on RISK also worked with the inclusion and exclusion of the data variables of ethnicity, gender and age, shaping citizens' data relation to the state with a view to fairness. While they argued that it would be unethical to ethnicity and gender in the model from the outset, the insight that age could perpetuate bias was prompted by a university student's bachelor thesis, communicated in the Danish magazine Zetland:

> She [the student, Therese Moreau] delved into the mathematics behind the calculations and began testing the algorithm. (…) [She] also noticed something mysterious: Age had a disproportionately large influence on the score. If a 2-year-old had been subjected to sexual abuse, the risk score was 7. For a 10-year-old, it was 9, and for a 12-year-old, it was 10. It suggested that the algorithm believed the severity of abuse increased with the child's age. Therese Moreau thought it couldn't be right. After all, age is just, well, age. It says nothing about the child's situation. There must be a mistake, she thought, a very serious one. (Kulager 2021)

In the article's comments section, the researchers acknowledged: 'Regarding age, it is entirely correct that there may be an issue we have not been aware of. We have investigated bias/fairness concerning gender and ethnicity but have not been attentive to age bias. That work has already been initiated concerning the new models we are in the process of developing' (in Kulager 2021). In a subsequent interview, they explained how this finding led to a change of the weight of age in the model, i.e., a reconfiguration of its afterlife:

> The consideration was whether by including age, we might assign too low a risk to the well-being of younger children because it is typically slightly older children who are placed in care. There may be valid reasons for this, but the risk is that younger children may go unnoticed. Therefore, we took the approach of conducting the decile classification within each age group. This means that in the decile with the highest risk per construction, there are 10% of the 0-year-olds, 10% of the 1-year-olds, 10% of the 2-year-olds, and so on. The same applies to all other deciles (SDS, December, 2023).

In this way, a non-affiliated bachelor student's discovery of an age bias led them to reflect on the relationship between outcome measure (children's maltreatment) and proxy (out-of-home placement). Although the algorithm discovered through machine learning that children, on average, are placed in out-of-home care arrangements at relatively high age, this, the researchers explained, does not necessarily mean that they do not experience distress earlier, but only that they are less likely to be removed. Thus, they overruled the model's initial age bias by distributing it equally among children aged 0–9 years. Here, we, thus, see how public involvement and critique led to changes in citizens' data afterlives, here by leading the researchers to remove the data trace left by the variable of age.

In examining bias on ethnicity and gender, they took a different approach. In contrast with the variable of age, they

did not include data on ethnicity and gender in the development of the algorithmic model. Yet, they did include it in a statistical analysis of fairness after the phase of model development. As they reflected,

> In practice, the municipalities have access to these data (…). But [we] estimated that it would not be an objective foundation for this tool (…) if all other risk factors were equal, it would not be appropriate if (…) the model assigned a boy with a higher risk score than a girl (JDS, Oct 2021).

At the same time, fairness was not simply a matter of excluding the variables of gender and ethnicity but rather an attribute of the model. Thus, they still saw the need to conduct a statistical fairness analysis as part of the post-validation of the algorithm, even if data points deemed to perpetuate biases [gender, ethnicity] had not been included. They estimated that biases could be contained in the algorithm even if variables conventionally assumed to reinforce biases were not included in its development.

> This means that we have a model that does not consider gender or ethnicity. However, we do, after the model has been build, examine whether the model's output holds biases with regards to the child's gender or ethnicity. (…) One could imagine that other data points in the model would correlate with some of the things [ethnicity and gender], which we did not include. So, we made these types of post-analyses [to investigate potential bias]. (…) (JDS, Oct 2021).

Referring to fairness literature in machine learning, they instead saw bias as a function of the model, of correlations between risk scores and variables of ethnicity and gender. In other words, to the data scientists, bias was best reduced by *adding* data on ethnicity and gender in post-fairness analyses, to examine correlations between the model's outputs and the data points of gender and ethnicity. Biases here are seen exist in the relationships in data, even if these data sets do not include the variables of gender and ethnicity. This amounts to a difference between bias as something pertaining to certain variables verses bias as a property of the model that can be mathematically reduced *ex post*.

The post-inclusion of these data sets for purposes of fairness, thus, functioned to further refine the relationships between data already contained in the model. E.g., if the researchers found an ethnicity bias in the model, they could identify the relationships producing that bias and try to reduce them in the model. This speaks to Tone Walford's (2013) ethnography of the production of scientific data in the Amazonas where she examines scientists efforts going in 'refining' and 'singularizing' relations in data sets. The inclusion or exclusion of data points was not simply a binary effort, but instead entailed modes of taking different factors into account 'in order to discount each one' (Walford 2013). For our case, the smaller data sets the researchers wanted to use as input data, the more data points they needed to post-validate the algorithm in terms of accuracy and fairness. Like Walford argues, 'the smaller the relation one is cutting out of the world, the bigger the world must become' (Walford 2013: 53). Interestingly, from a machine learning perspective, then, bias could be reduced through the *inclusion* (*ex post*) of datasets (gender and ethnicity) that from a legal and ethical perspective needed to be omitted from the model.

6 Concluding remarks

Public sector adoption of machine learning techniques raises important questions about state–citizen data relationships. In this paper, we have made inquiries into citizen data afterlives, examining how decisions are made about which datasets to include in a research project using machine learning to develop and evaluate a model for child protection services. Data afterlives reflect how administrative registers, containing data about citizens' past interactions with the public authorities, become 'assets' (cf. Birch *et al* 2021) in the endeavor to develop new techniques for scoring, targeting, and profiling citizens. Repurposing national archives, such as the archive of referral, for the development of predictive models, implies a temporal shift, from one of learning from the past through comparison to predicting uncertain futures through machine learning. Thus, rather than coming to stand for municipal failure (e.g., failing to react on a referral, as hinted by the minister), as data afterlives, referrals come to stand for anticipation and pre-emption of unwanted futures, here at the level of the individual rather than population as with register research (cf. Cevolini and Esposito 2020). This renders the question of how citizens' interactions with the State give rise to new data afterlives particularly salient to study.

Our study identifies three types of citizen data afterlives: ground truth, post-validation, and fairness. Each of these serves a particular purpose in the research project and each is shaped by frictions and negotiations across researchers, forms of expertise and public critique. Here we found a bifurcation of citizens' data afterlives in data sets for, respectively, development (ground truthing, training, and testing the algorithmic model) and evaluation beyond that entailed in the machine learning process (post-validation and fairness analysis). Citizens' data afterlives figure differently in these two forms. With machine learning, used to identify patterns and relationships within the data and turn this into an algorithmic model, the role of citizen data is to infer futures, i.e., made predictions about citizens' unknown futures. As the algorithmic model gained capacity to generate and infer rules from the examples in data, these data

sets also came to stand for a promise of making inductive and inferential forms of knowing, classifying, and deciding. In post-validation and fairness analyses, in turn, citizens' data afterlives took a different form of inclusion. Rather than inferring futures, data sets here functioned to test, validate and tweak the model with a view to rendering it legal, precise, fair, and ultimately, more trustworthy. Here, data sets, thus, worked to establish the model's accuracy and fairness, qualities potentially adding to its legitimacy. Interestingly, the research project's focus on data minimization and ethics for model development (i.e., training and testing) generated new types of 'shadow' data afterlives, where datasets were used in the research process to interrogate the validity, legality, and fairness of the model rather as opposed to expanding the model's capacity for surveillance.

Studying machine learning model development in a Danish welfare context provides an interesting contrast to 'big data' regimes of data consumption. In our case, we do not see a 'relentless' hunger for 'big data' and much care is taken to limit the model as much as possible, both as part of their intended ethical reflections (e.g., excluding data on ethnicity and gender) and unexpected public engagement (resulting in the discovery of age bias). Thus, we add to Reutter's (2022) finding that the Scandinavian welfare context poses institutional and regulatory limitations for datafication of public service delivery. We suggest that careful attention to practices of dataset inclusion, as well as eliciting the types of afterlives that do not aim at surveillance or prediction, gives a more nuanced picture of data afterlives in Scandinavian algorithmic and predictive modes of governance. These, we suggest, cannot be reduced to American big tech surveillance logics but form a more complex image. Not at least in a (Danish) situation where many attempts to develop predictive algorithmic models for public administration are canceled due to issues with legality, data quality, and digital infrastructures (Ratner and Schrøder 2023). Albeit crucial concerns with regards to citizens' basic rights in algorithmic regimes remain (Akhtar and Jørgensen 2021), projects such as RISK may also indicate a different politics of 'data justice' (Grant 2020) where a local context of public concern, institutional learning, and legal regulation call for analytical approaches that can take the socio-political context of data afterlives into account.

The contemporary and unsettled negotiations about whether citizens' data afterlives, stored in national archives, should be repurposed for predictive models, are important. Our study shows how the development of such models entail many different types of data afterlives that do not simply reproduce archival representation on a 1:1 basis. While obscure to the public eye, examining such algorithmic arrangements are nevertheless important for thinking through the ethico-politics of citizen–state relations. By highlighting how many different concerns are negotiated in decisions about dataset inclusion, we, thus, not only offer insight into the social practices that shape a particular model, but also provide material for further debates about how specific contexts of model development, such as regulation and public debate in a context of the Danish welfare state, shape citizens' data afterlives in machine learning projects. This offers crucial insight into the 'plural branching pathways that could have yielded a different output and to amplify those branches as political decisions' since 'in every arrangement of a machine learning model there are the traces of the rejected alternative' (Amoore 2023: 35). As this study shows, those developing the model are painfully aware of these branching points; our hopes with this article is that they are translated into the wider public too.

Acknowledgements We express our sincere gratitude to the interlocutors at RISK for their generosity in taking their time to share insights, perspectives and thoughts as well as patiently explaining their practices. Their contributions have enriched the scope and depth of our research. Additionally, we extend our heartfelt appreciation to the organisers and participants of "Reframing ADM: Concepts, Values, Alternatives" (29-30 Aug. 2022) and "Welfare After Digitalization" (28-29 Nov. 2022) for feedback on earlier versions of the article. Moreover, we would like to acknowledge the reviewers for their constructive and thoughtful feedback and the editor for their guidance and support throughout the publication process. And finally, we extend our gratitude to the the VELUX FOUNDATIONS (Algorithms, Data & Democracy Jubilee Grant) and IRFD (9131-00115B: AI Reuse and 0132-00080B: Datafied Living) for supporting this research.

Curmudgeon Corner Curmudgeon Corner is a short opinionated column on trends in technology, arts, science and society, commenting on issues of concern to the research community and wider society. Whilst the drive for super-human intelligence promotes potential benefits to wider society, it also raises deep concerns of existential risk, thereby highlighting the need for an ongoing conversation between technology and society. At the core of Curmudgeon concern is the question: What is it to be human in the age of the AI machine? -Editor.

Funding Open access funding provided by Aarhus Universitet.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest. Data cannot be made available due to reasons of confidentiality and ethics.

References

Agostinho D (2019) Archival encounters: rethinking access and care in digital colonial archives. Arch Sci 19(2):141–165

Akhtar, M, Jørgensen, RF (2021) Når algoritmer sagsbehandler – Rettigheder og retssikkerhed i offentlige myndigheders brug af profileringsmodeller. Danish Institute for Human Rights.

Amoore L (2020) Cloud ethics: algorithms and the attributes of ourselves and others. Duke University Press, Durham, NC

Amoore L (2023) Machine learning political orders. Rev Int Stud 49(1):20–36

Birch K, Cochrane D, Ward C (2021) Data as asset? The measurement, governance, and valuation of digital personal data by Big Tech. Big Data Soc 8(1):20539517211017308

Cevolini, A, Esposito, E (2020) From pool to profile: Social consequences of algorithmic prediction in insurance. Big Data & Society 7(2). https://doi.org/10.1177/2053951720939228

Dencik L, Redden J, Hintz A, Warne H (2019) The 'golden view': data-driven governance in the scoring society. Internet Policy Rev 8(2):1–24

Denton E, Hanna A, Amironesei R, Smart A, Nicole H, Scheuerman MK (2020) Bringing the people back in: contesting benchmark machine learning data sets. arXiv

Ebeling MFE (2022) Afterlives of data. Life and debt under capitalist surveillance. University of California Press, Oakland, CA

Edwards PN, Mayernik MS, Batcheller AL, Bowker GC, Borgman CL (2011) Science friction: data, metadata, and collaboration. Soc Stud Sci 41(5):667–690

Elleman K (2015) Minister: underretningsstatistik om udsatte børn og unge nyt vigtigt redskab for kommunerne. Danish Ministry of Interior and Health. https://im.dk/nyheder/nyhedsarkiv/2015/nov/minister-underretningsstatistik-om-udsatte-boern-og-unge-nyt-vigtigt-redskab-for-kommunerne

Erlangsen A, Fedyszyn, I (2015) Danish nationwide registers for public health and health-related research. Scand J Public Health 43:333–339. https://doi.org/10.1177/1403494815575193

Eubanks V (2018) Automating inequality: how high-tech tools profile, police, and punish the poor. St. Martin's Press, New York

Frederiksen M (2020) Prime Minister's new year speech, Jan 1st. https://www.altinget.dk/artikel/mette-frederiksens-nytaarstale-flere-udsatte-boern-skal-have-et-nyt-hjem

Grant A (2020) Predictions, Mocks or Models? Learning from cancelled predictive analytics in public services. Carnegie UK Trust. https://carnegieuktrust.medium.com/predictions-mocks-or-models-learning-fromcancelled-predictive-analytics-in-public-services-e6bba658c130. accessed 14 Aug 2022

Hacking I (1991) The making and molding of child abuse. Crit Inq 17(2):253–288

Hanna A, Denton E, Amironesi R, Smart A, Nicole H (2020) Lines of sight. Logic Magazine. https://logicmag.io/commons/lines-of-sight/

Hansen KB, Borch C (2022) Alternative data and sentiment analysis: prospecting non-standard data in machine learning-driven finance. Big Data Soc 9(1):1–14. https://doi.org/10.1177/20539517211070701

Hartley JM and Thylstrup, NB (2024) The Algorithmic Gut Feeling–Articulating Journalistic Doxa and Emerging Epistemic Frictions in AI-Driven Data Work. Digital Journalism, 1–20.

Heuts F, Mol A (2013) What is a good tomato? A case of valuing in practice. Valuat Stud 1(2):125–146. https://doi.org/10.3384/vs.2001-5992.1312125

Hoeyer K (2019) Data as promise: reconfiguring Danish public health through personalized medicine. Soc Stud Sci 49(4):531–555. https://doi.org/10.1177/0306312719858697

Hoeyer K (2023) Data paradoxes: the politics of intensified data sourcing in contemporary healthcare. MIT Press, Cambridge

Jaton F (2017) We get the algorithms of our ground truths: Designing referential databases in digital image processing. Soc Stud Sci 47:811–840.https://doi.org/10.1177/0306312717730428

Jaton F (2021) The constitution of algorithms: ground-truthing, programming, formulating. MIT Press, Massachusetts

Jo ES, Gebru T (2020) Lessons from archives: strategies for collecting sociocultural data in machine learning. In: Proceedings of the 2020 conference on fairness, accountability, and transparency. pp 306–316

Jørgensen RF (2023) Data and rights in the digital welfare state: the case of Denmark. Inf Commun Soc 26(1):123–138. https://doi.org/10.1080/1369118X.2021.1934069

Jucan IB, Parikka J, Schneider R (2019) Remain. U of Minnesota Press, Minneapolis

Kaufmann M, Leese M (2021) Information in-formation: algorithmic policing and the life of data. In: Završnik A, Badalič V (eds) Automating crime prevention, surveillance, and military operations. Springer, Cham, pp 69–83. https://doi.org/10.1007/978-3-030-73276-9_4

Keenan T (2018) Getting the dead to tell me what happened: Justice, prosopopoeia, and forensic afterlives. Kronos, 44(1):102–122.

Kristensen K (2022) Hvorfor Gladsaxemodellen fejlede—Om anvendelse af algoritmer på socialt udsatte børn. Samfundslederskab i Skandinavien 37(1):27–49. https://doi.org/10.22439/sis.v37i1.6542

Kulager F (2021) Kan algoritmer se ind i et barns fremtid? I Hjørring og Silkeborg eksperimenterede man påudsatte børn. Zetland. https://www.zetland.dk/historie/s8YxAamr-aOZj67pz-e30df. Accessed 15 Jan 2023

Lee F, Helgesson C-F (2020) Styles of valuation: algorithms and agency in high-throughput bioscience. Sci Technol Human Values 45(4):659–685. https://doi.org/10.1177/0162243919866898

Leonelli S, Tempini N (2020) Data journeys in the sciences. Springer, Cham

Leslie D, Holmes D, Hitrova C, Ott E (2020) Ethics review of machine learning in children's social care. What works for children's social care. http://whatworks-csc.org.uk/research-report/ethics-review-of-machine-learning-in-childrens-social-care/

Mackinnon K (2022) Critical care for the early web: ethical digital methods for archived youth data. Journal of Information, Communication and Ethics in Society, 20(3):349-361.

Medina Perea IA, Cox A, Bates J (2020) Exploring the life of patient data in the UK healthcare sector. AoIR Selected Papers of Internet Research. https://spir.aoir.org/ojs/index.php/spir/article/view/11279

Nadim T (2016) Data labours: how the sequence databases GenBank and EMBL-Bank make data. Sci Cult 25(4):496–519

Odumosu T (2020) The crying child: On colonial archives, digitization, and ethics of care in the cultural commons. Current Anthropology, 61(S22):289-302.

Paullada A, Raji ID, Bender EM, Denton E, Hanna A (2020) Data and its (dis)contents: a survey of data set development and use in machine learning research. arXiv Preprint arXiv:2012.05345

Plantin JC (2019) Data cleaners for pristine data sets: visibility and invisibility of data processors in social science. Sci Technol Hum Values 44(1):52–73. https://doi.org/10.1177/0162243918781268

Plesner U, Justesen L (2022) The double darkness of digitalization: shaping digital-ready legislation to reshape the conditions for public-sector digitalization. Sci Technol Hum Values 47(1):146–173. https://doi.org/10.1177/0162243921999715

Raji ID, Buolamwini J (2022) Actionable auditing revisited: investigating the impact of publicly naming biased performance results of commercial AI products. Commun ACM 66(1):101–108

Raji ID, Bender EM, Paullada A, Denton E, Hanna A (2021) AI and the everything in the whole wide world benchmark. arXiv Preprint arXiv:2111.15366

Ranchordas S (2021) Empathy in the digital administrative state. Duke Law J (Forthcoming), University of Groningen Faculty of Law Research Paper No. 13/2021, 1–45. https://doi.org/10.2139/ssrn.3946487

Ratner HF, Elmholdt KT (2023) Algorithmic constructions of risk: anticipating uncertain futures in child protection services. Big Data Soc 10(2):1–12. https://doi.org/10.1177/20539517231186120

Ratner HF, Ruppert E (2019) Producing and projecting data: aesthetic practices of government data portals. Big Data Soc 6(2):1–16. https://doi.org/10.1177/2053951719853316'

Ratner HF, Schrøder I (2023) Ethical plateaus in Danish child protection services: the rise and demise of algorithmic models. Sci Technol Stud XX(X): 1–18. https://doi.org/10.23987/sts.126011

Redden J, Dencik L, Warne H (2020) Datafied child welfare services: unpacking politics, economics and power. Policy Stud 41(5):507–526. https://doi.org/10.1080/01442872.2020.1724928

Reutter L (2022) Constraining context: Situating datafication in public administration. New Media & Society 24:903–921. https://doi.org/10.1177/14614448221079029

Ribes D (2017) Notes on the concept of data interoperability: cases from an ecology of AIDS research infrastructures. In: Proceedings of the ACM conference on computer supported cooperative work, CSCW https://doi.org/10.1145/2998181.2998344

Ribes D, Hoffman AS, Slota SC, Bowker GC (2019) The logic of domains. Soc Stud Sci 49(3):281–309. https://doi.org/10.1177/0306312719849709

Scheuerman MK, Hanna A, Denton E (2021) Do data sets have politics? Disciplinary values in computer vision data set development. Proc ACM Hum Comput Interact 5(CSCW2). https://doi.org/10.1145/3476058

Schneider R (2011) Performing remains: art and war in times of theatrical reenactment. Taylor & Francis, New York

Slota SC, Hoffman AS, Ribes D, Bowker GC (2020) Prospecting (in) the data sciences. Big Data Soc 7(1):2053951720906849. https://doi.org/10.1177/2053951720906849

Sutherland T (2023) Resurrecting the black body: race and the digital afterlife. University of California Press, Berkeley, CA

Thylstrup NB (2022) The ethics and politics of data sets in the age of machine learning: Deleting traces and encountering remains. Media, Culture & Society, 44(4):655-671.

Villumsen AM, Søbjerg LM (2020) Informal pathways as a response to limitations in formal categorization of referrals in child and family welfare. Nordic Soc Work Res 13(2):176–187. https://doi.org/10.1080/2156857X.2020.1795705

Walford A (2013) Transforming data: an ethnography of scientific data from the Brazilian Amazon. IT University of Copenhagen, Copenhagen

Winthereik BR (2023) Data as relation: ontological troubles in the data-driven public administration. Comput Supported Coop Work.https://doi.org/10.1007/s10606-023-09480-9