Ethical Plateaus in Danish Child Protection Services: The Rise and Demise of Algorithmic Models

Helene Friis Ratner

Danish School of education, Aarhus University, Denmark/ helr@edu.au.dk

Ida Schrøder

Aarhus University, Denmark

Abstract

This paper analyses how controversies shape an emerging field of AI in Danish child protection services. In a context of high controversiality, we examine how algorithmic systems evolve in conjunction with changing ethical stakes. Empirically, we report a study comprising all Danish attempts (*n=4*) to develop algorithmic models for child protection services. These attempts were never fully implemented and have been either cancelled, paused or changed significantly since their outset. Combining the notion of 'ethical plateaus' with insights from valuation studies, we propose that public controversies shape how organisations enact their algorithms as ethically 'good'. Our findings demonstrate how valuations of ethically contestable algorithms involve the very distribution of agency across humans and algorithms, i.e., how much power and agency should be delegated to algorithmic models. In the case of Danish child protection services, this moves towards reducing their agency.

Keywords: Artificial intelligence, Ethics, Controversy, Child Protection Services

Introduction

Heralded with promises of increased scope, speed, and precision, algorithmic systems based on machine learning are making their way into child protection services (Andrejevic, 2017; Meilvang and Dahler, 2022; Redden et al., 2020). At the same time, algorithmic technologies have spurred an avalanche of ethical concerns regarding the powers of algorithms (Beer, 2009; Mittelstadt et al., 2016). Such concerns do not only exist in academic literature but are increasingly put on public display through medialised controversies (Kristensen, 2022). However, as Noortje Marres (2021) argues, public controversies are not simply democratic contestations of emerging technologies but increasingly figure as trial grounds for innovation and product development. Addressing such an interplay between medialised controversy and the development or closure of algorithmic models, we are interested in its implications for changing ethical boundaries of AI[1] in child protection services.

Internationally, policy makers and managers have begun to investigate predictive algorithms as tools to support social workers' assessments and decision-making. Yet, the critiques of the ethi-

cality of algorithmic models in child protection services abound. Rather than help children, the critiques go, algorithmic models instead result in biased and non-transparent decisions, increased inequality, and a de-humanised administration based on datafied surveillance of the poor (Dencik et al., 2019; Eubanks, 2018; Jørgensen, 2021). The high level of controversy in the field of child protection renders the question of ethics especially prevalent and makes it a good setting for exploring empirically how the ethics and value of algorithms are developed and change in conjunction with public value contestation and criticism.

We propose to analyse this emerging and changing field of 'ethical AI' in child protection through the concept of 'ethical plateaus' (Fortun and Fortun, 2005; Seaver, 2021). This concept highlights the instability of an ethical terrain with moving boundaries of what is ethically possible. To examine how organisations negotiate the value and ethics of their algorithmic model we further draw on insights from valuation studies (Helgesson and Muniesa, 2013). Rather than assuming 'value' to be an inherent property of an object, valuation studies emphasise value as *enacted*, i.e., as a practical accomplishment and attribution. Together, these concepts offer useful sensibilities for the analysis of how the ethical stakes of publicly contested algorithmic systems change over time and how this impedes the very valuation of algorithmic models. Thus, in examining how public controversy, ethical stakes and valuations of algorithmic models shape one another in 'changing ethical plateaus', we ask how algorithmic models are enacted as valuable, how they are rendered controversial and why they, eventually, were cancelled or changed.

Empirically, we report a qualitative study of four attempts to develop algorithmic models for public sector child protection agencies in Denmark. The four algorithmic models comprise all Danish attempts to design, develop, test, and implement algorithmic models in this area of work and they took place during a six-year period (2017-2022). Although predictive algorithms used to profile children and families at risk are gaining ground internationally, especially in anglo-saxon countries (Cuccaro-Alamin et al., 2017; Redden, 2018; Russell, 2015), these systems are deemed highly controversial in Denmark, and none of the four algorithmic models have been implemented. Indeed, the examples we examine in this paper never made it beyond the status of project designs, limited trials, or research. This is curious for a country, which is otherwise ranked as the number one country in digital government across the world (United Nations Department of Social and Economic Affairs, 2022). In this sense, the Danish case provides a rather unique entry point into understanding how ethical plateaus emerge and undergo change in conjunction with organisational attempts at developing and rendering algorithmic models in child welfare valuable and legitimate. Furthermore, studying algorithms that have been cancelled or changed significantly provides a fruitful entry point into valuation practices as the actors developing the algorithms, first, have been called to explicate how and why their (envisioned) algorithm is ethically good, and second, justify or explain why they had to close or change it.

The paper is structured as follows. Taking our inspiration from anthropology and Science and Technology Studies (STS), we provide a brief review of some important discussions here. This is followed by a presentation of our conceptual approach, the combination of 'ethical plateaus' and 'valuation', as well as a methods section, introducing the context of child protection services and accounting for the empirical data assembled and analysed for this paper. In the subsequent analytical sections, we present our analysis of the four Danish cases, identifying how and why the different organisations embarked on using AI for child protection, how valuations changed over time, sometimes in response to public controversies, as well the events that led to cancelling, pausing or changing the algorithmic projects. In a concluding discussion, we develop a timeline across the four cases, visualising how ethical plateaus change over time and how public controversies and valuations intersect. Over time, the algorithmic models are granted less agency and power. This indicates that ethical plateaus are closely entwined with negotiations of how agency and power should be distributed across humans and machines, figuring the politics of what it might mean to live with AI and what we mean by ethical AI.

Ethics as socio-material and processual accomplishments

This paper builds on the work of scholars associated with STS and anthropology who are beginning to examine ethics as a processual, relational and practical accomplishment. Puig de la Bellacasa, for instance, conceptualises ethics as "concrete [socio-material] relationalities in the making" rather than "normative morals" (Puig de la Bellacasa, 2010: 152). From this follows that we should not presume knowing in advance what ethical AI should look like, but instead, examine its ongoing, uncertain and situated making. Many scholars in STS have followed this call. In his study of music recommender algorithms, Seaver (2021), for instance, explores how the makers of recommender systems engage with popular critiques of algorithms as powerful, computerised agencies replacing 'careful human judgment' of music. Seaver (2021: 512) analyses developers' reasoning of these critiques as the business of actively "making ethics, trying to understand, evaluate, and reconfigure the field of possible choices". Ethics, in this view, are enacted in the developers' framing of how competing values can co-exist.

Another example is Douglas-Jones (2017) who has conducted an ethnographic study of the ethics review committees for biomedical research in the Asia-Pacific region. Exploring ethics as a material practice, she examines how universal ethical standards are negotiated in encounters with situated and socio-material circumstances such as office spaces. This focus allows her to study the mundane work of building infrastructures for ethical review and universal standards, concluding that universal ethics emerge "as a site of ongoing attention and negotiation, standard making and aspiration" (Douglas-Jones, 2017: 28). Finally, Ziewitz (2019) has studied ethics as a practical accomplishment in a context of SEO consultants work with search machine optimisation, aiming to understand 'how people organise themselves as ethical in the absence of the ontological security that professional ethicists and some philosophers presume'. Like Puig de la Bellacasa (2010), Ziewitz emphasises the need to leave behind the inclination to decide whether a practice is ethically correct and instead accomplish a deeper understanding of how "'being ethical' [is] (…) implicated in and organize the (…) experiences of people" (Ziewitz, 2019: 713). This approach, thus, allows a close look into the practical work of establishing, negotiating and distributing ethics in a context of its increasing contestation.

In our analysis of how the valuation of algorithms change in conjunction with changing ethical stakes, we build on these sensitivities. Examining the interplay between medialised controversies and algorithmic projects in Danish child protection services, we contribute with an inclusion of public controversy as an important factor in the development of what counts as ethical AI in child protection.

Ethical plateaus and valuation studies

The notion of 'ethical plateaus' helps us conceptualise changing ethical boundaries of what is possible in techno-scientific situations fraught with dilemmas and ethical contestations. Defining ethical plateaus as a site "where multiple technologies interact to create a complex terrain or topology of perception and decision making" (Fischer in Fortun and Fortun, 2005: 47), the concept allows one to examine the intersection and co-evolvement of different ethical concerns. For our purpose, we do this across four Danish algorithmic models in child protection services, attending to how the developers attempt to manage the horizons of possible ethical issues posed by controversial algorithms and how the different projects relate to one another, e.g., through practices of 'un-ethicizing' (Tønnesen, 2009).

The concept's geological metaphor brings about the image of a complex socio-technical landscape made up by interactions between, in our case, algorithmic models, administrative apparatuses and public media controversies. Like geological plateaus, formed through processes such as volcanic activity, tectonic uplift, or erosion, ethical plateaus constitute a dynamic terrain of changing and competing concerns and values.[2] In this context, we might see public controversies as volcanic 'ruptures' shaping the formation of plateaus, insofar as they make organisations re-value and modify their algorithmic models significantly. The ethical plateaus, then, "foregrounds the tectonics of the ethical sphere—everything that supports and constrains the

range of ethical possibility, without making a strong distinction between the 'hard' constraints typically associated with technology and the 'soft' ones associated with society" (Seaver, 2021: 513). Following from this, we are not interested in assessing the ethicality of the algorithmic models we study by benchmarking them against established ethical principles. Instead, we explore how valuations of the ethicality of algorithms change over time in conjunction with public controversy, understanding this process as changing ethical plateaus and their ongoing figurations of what is ethically acceptable to do with AI in Danish child protection services.

As part of this, we draw on the insights from Valuation Studies, allowing us to depart from the idea that the value of an algorithm is an inherent attribute or quality of the algorithm itself. Instead, this approach emphasizes value to be the result of a situated and practical endeavour to explicate what the algorithm is good for (Helgesson and Muniesa, 2013). I.e., if we are to learn what is valuable about an algorithm, we must look for the situated valuations of algorithms. In this view, value conflicts do not occur between pre-established ideas of what is good and valuable in a society, but rather as practices of negotiating, adjusting, and reconceptualising the algorithms. Analysing changing valuations of what counts as the ethically good algorithm in conjunction with public controversies allow us to draw the contours of emerging ethical plateaus of AI in Danish child protection services.

Empirical resources

Our study comprises of all (four) Danish attempts to develop algorithmic models for child protection: Three municipal development projects and one research project named RISK ("Underretninger i fokus"). RISK and one of the municipalities, Gladsaxe, were subject to public controversy. As Gladsaxe and RISK have been under much public scrutiny, pseudonymization is impossible here. The other two municipalities are pseudonymised.

For all cases, we conducted interviews with key stakeholders (e.g., project leaders, data scientists, municipal directors) amounting to 39 interviews. We collected 63 public- and non-public documentations (e.g., project descriptions, minutes, legal assessments, power point presentations) and traced public media-debates concerning the roles of algorithms in Danish public administration (source: Infomedia), including 45 media articles. For all interviews, we had our interviewees make a timeline and identify important turning points. For those projects subject to public media controversy, we also asked our interlocutors to reflect on selected critical media quotes.

In analysing the empirical material, we developed timelines for all cases, identifying 1) the initial justifications for developing the algorithmic models and initial ethical considerations, 2) public and non-public contestations and critiques of their ethicality as well as project members' responses to these, and 3) why they were eventually closed or changed considerably. For these moments, we analysed valuations. i.e., their enactment of the (ethical) 'goodness' of the algorithmic models. We further made note when project agents explicitly related their own algorithmic model to the other cases, to examine how they made sense of their own algorithm in comparison to the other Danish examples. As our aim is to identify emerging ethical plateaus, we have chosen the valuations that we deemed to be most dominant and influential in the development (and termination) of the algorithms. Our analysis thus does not reflect all valuations detected in our mapping and, thus, we make no claims of completeness.

The context of Danish child protection services

Child protection services are tasked with the difficult but crucial task of preventing and stopping child maltreatment. Identifying children at risk of maltreatment, however, is a complex task, fraught with uncertainty and severe consequences for the families involved if the wrong assessments are made (Villumsen and Søbjerg, 2020). Even in Denmark, where the universalistic welfare model entices professional collaborations across core welfare services such as daycare, healthcare, and school, and where the Social Service Act demands early interventions with family-oriented services, four percent of children are at some point during their childhood placed in out of home institutions or foster families due to neglect, maltreatment

or other situations which impedes on the child's wellbeing and abilities to develop alongside with his or her peers (VIVE, 2022).

How to break these statistics has been a pivotal concern in the past 20 years of reformation of the Danish child welfare system. Most importantly the welfare professionals' obligation, as well as private citizens' access, to notify the authorities about children they are concerned about was immensely expanded from 2011 and onwards. Correspondingly, the number of notifications about children failing to thrive has increased ever since – from 97.288 notifications in 2015 to 138.099 notifications in 2021 (Statistics Denmark, 2022). Combined with municipalities' legal demand to assess the severity of notifications within 24 hours upon receival, the system of notifications has become pivotal in the organising and innovation of Danish Child protection agencies. As we will see in the following, the four algorithmic projects entail rather different 'problematisations' (cf. Callon, 1986) of this situation, two (the municipalities) targeting the massive amounts of notifications, one (RISK) targeting the assessment of these notifications and, finally, one (Gladsaxe) trying to pre-empt notifications through early intervention (cf. Ratner and Elmholdt, 2023).

The Gladsaxe model: Algorithmic prediction to pre-empt unwanted futures

The first attempt to develop a predictive algorithm in the field of child protection was undertaking by Gladsaxe municipality in 2017-2018. In 2017, the civil servants and politicians of Gladsaxe municipality formulated the idea of developing the algorithmic tool for "data-driven early detection". This model was to solve a problem of being notified of children's problems too late. In Denmark, child protection services learn about children's maltreatment through notifications sent to them, i.e., concerns about children's wellbeing. As the then leader of the child protection services explained, it was not simply a problem of welfare professionals failing to notify them but a problem of the municipality not linking up data already held by different welfare departments:

> I mean, how do we reach these families when their children are infants? (…) At some point, our leader of the employment services remarks that they are the first to be advised when a long-term unemployed mother is pregnant. (…) But the problem is that this information stays with the employment services because there is no 'forward information-button', you know? We are not notified in the child protection services so we can't begin working with these mothers. (…) As it is now, we are simply waiting for the children's symptoms – that something is wrong in the family – to emerge, instead of acting on our knowledge of the risks being present in a family. (Interview, June 2021, leader of Child protection services)

Illustrating how long-term unemployment is considered a 'risk', the leader reflects on how such data on risk resides with other welfare departments but rarely reaches the child protection officers. This realization made them think about how to bypass notifications being sent 'too late'. An algorithmic merging of data from different welfare departments, the idea was, could serve as an alternative mode of detecting children *before* symptoms would emerge. The algorithm was envisioned to merge data on known risks such as parents' employment status and history, substance abuse, absence from appointments with the dentist or health nurse, earlier notifications to child protection services. The idea was to use the algorithm for detection of at-risk families, after which a case worker would make contact and offer help on a voluntary basis. This valuation of the goodness of the algorithm relies on a distinction between risk (here attributed to e.g. parents with long-term unemployment) and symptoms. Rather than acting on 'symptoms' on children's maltreatment, of which they are notified in notifications, the algorithm is valued for its capacity to detect risks and thereby pre-empt children's possible maltreatment through anticipation.

In valuing the algorithm as ethically good, the developers further emphasised that only the computer would access citizens' data. A municipal leader said it like this in an interview:

> I mean, all these data would be in a black box which neither I nor other employees could access. (…) We don't need to see all these [the municipality's corpus of] citizens' data. We are not

> interested in this. They [the citizens] need to live their lives out there. It is just the small share [of children] which we believe we can pull out (…). And these are the ones we want the algorithm to find. (Interview, June 2021, leader of Child protection services)

With this description of the algorithm, they established a distinction between the computational analysis of all citizens' data and the caseworkers' access to the select citizens profiled by the algorithm. They visualised the ethicality of the algorithm as a 'statistical black hole', emulating the algorithm's analysis of citizens' data as non-visible to employees (figure 1), demarcating this as more ethical than case workers looking through all this data.

This valuation invokes the employees' non-access to citizens' data as an ethically valuable property of the algorithmic model. It also enacts surveillance as a human endeavor, i.e., algorithmic profiling would not count as surveillance unless a human caseworker is being informed of the identity of the profiled citizen.

For the algorithm to be tested and implemented, the municipality applied to the Ministry of Interior to be exempted from privacy regulation requiring citizens' consent to merge data. To the municipality's' surprise, the Ministry of Interior rejected their application with the argument that they really liked their idea and wanted to propose the government to change to this legislation rather than granting one municipality exemption. As the data scientist noted: "I mean, it really surprised us that they didn't want us to test it [in just our municipality] before granting all municipalities legal permission to do this" (interview, June 2021). At this point, the emerging ethical plateaus of AI in child protection, are thus rather wide. The algorithmic model is valued for enabling the "early detection of (…) risk factors in the parents before symptoms of maltreatment appear with the child, and hereby secure an earlier and more effective prevention of vulnerability" (Internal document[3]).

Controversy I: Algorithmic surveillance (2018-2019)

In March 2018, the Danish Government referred explicitly to Gladsaxe's application in their policy initiative to combat so-called 'ghettos'. One of the initiatives was to detect minority children assumed to be in extra need of protection from "parents [who] are affiliated with countries with other parenting traditions where violence is legal" and the algorithm was mobilized as a tool to identify these children (Danish Govern-

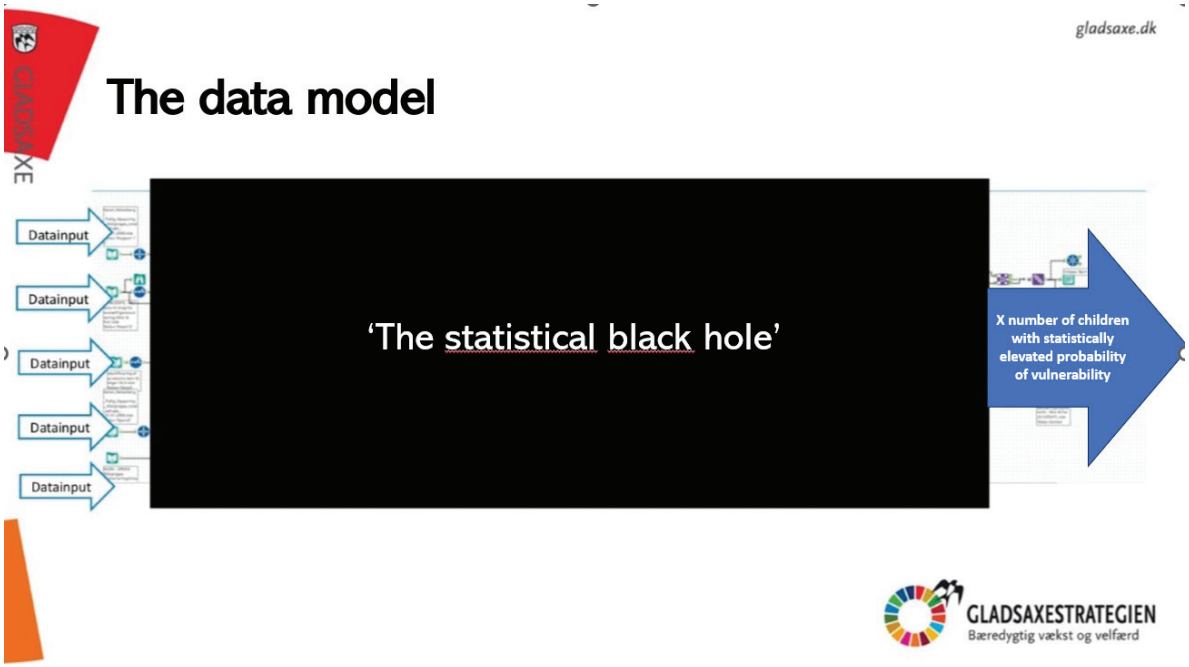

Figure 1. The statistical black hole, from Gladsaxe municipality power point presentation, 2018

ment, 2018: 30). This racialised figuration of the algorithm resulted in massive media attention to Gladsaxe's detection model. With headlines about "data-surveillance of families with children", a data-ethical controversy articulated issues of surveillance, biased decisions, a lack of transparency, risk of data misuse (Kjær, 2018). This public scandal introduced an important rupture to the emerging ethical plateaus of AI in child protection services, narrowing its wide planes contoured by techno-optimism to one establishing Gladsaxe's algorithmic model as unethical. As a result of the public controversy, several political parties withdrew their support to the profiling initiative in the Government's 'ghetto plan' and Gladsaxe had neither an exemption nor general legislation to support their tool (cf. Kristensen, 2022). With no legal mandate, the municipality had to put the algorithm on indefinite hibernate. Thus, the project was terminated before the algorithm could be fully developed and implemented.

The Gladsaxe model, and its racialised offspring as a 'ghetto plan', narrowed the ethical boundaries of what was possible to do with AI for child protection considerably. In this sense, Gladsaxe's project is also pivotal in the emerging ethical plateaus of AI for child protection services. Indeed, a search in Infomedia, a Danish newspaper- and magazine archive, revealed no hits on the terms child* AND algorithm* prior to 2018, whereas the number of articles continuously rose thereafter, most of them critical and focusing on Gladsaxe. This suggests that, at least in the media generated public, the coining of algorithms and child protection was not a matter of public attention prior to 2018. Today, the 'Gladsaxe model' has become a common point of reference in public and informal conversations about 'what can go wrong' in the use of predictive algorithms, to the extent that it has even become a hashtag (#Gladsaxemodel) in Twitter debates (the social media now known as X). Thus, we view the Gladsaxe experiment, the first attempt to develop AI for child protection services in Denmark, as the early formations of ethical plateaus in this field.

RISK (II): Algorithmic decision-support model to improve case workers' risk assessments

During the same period as the Gladsaxe experiment, a group of interdisciplinary researchers embarked to develop different predictive algorithm, here for the purpose of examining whether a decision-support tool could help social workers assess notifications. The potential value of this algorithmic model was envisioned in terms of hindering child maltreatment. As they wrote in a draft research article, "Child maltreatment has significant costs to its victims and, more generally, to society. Unfortunately, identifying cases of child maltreatment is a difficult task for Child Protective Services" (internal document). This difficulty was elaborated in their project description where they highlight a context of a growing number of notifications (from 97.288 in 2015 to 137.986 in 2019) (Project description: 2). The decision-support model was thus made valuable as a potential to help children but also a as research project assessing the efficiency of such a tool. Although sharing with Gladsaxe the objective of improving child protection, RISK's model also differs in important ways (Ratner and Elmholdt, 2023). First, rather than predicting risk *before* symptoms occur, RISK aimed predicting risk *after* symptoms had been notified. Second, rather than merging data from different welfare areas, they only wanted to use data that social workers could already legally access. Third, given their emphasis on the need to also research the value of such a tool from a social work and family perspective, their valuation of the algorithmic model as helpful was deliberately kept as an open question to be explored through research.

In the wake of the first rupture (I) in 2018, RISK published several reports on various aspects of their research endeavour, amongst them a report on their 'ethical considerations', dated October 2018. They sum up the ethicality of RISK, including testing it in social worker's practice, using the following words:

> It is ethically sound to test the tool in practice. Every day, assessments of notifications are made. The judgments and decisions, based on the assessments, are complex and they entail vast amounts of information and a series of ethical

> dilemmas, regardless of whether a statistical tool is being used or not. The ambition is to help children and young persons at risk in the best possible way. We adjudicate that the implementation of a statistical tool as a support for the qualitative assessment to be ethically sound and potentially improving the protection of vulnerable children and young persons" (Report on ethical considerations, October 2018).

Drawing the distinction between the complex, qualitative human assessment and the quantitative assessment offered by the algorithm, the role of the algorithm as a supplementary tool is emphasised. Here, the algorithm is valuable as a support in an already complex decision-making situation. With this valuation, the predictive algorithm is enacted as 'just' a tool, which will not render ethical dilemmas mores complex – because they are already complex. As a tool, in contrast, it might have the ability to improve how children and young persons are helped. Thus, the algorithmic processing of data was emphasised as ethically valuable, with the complexity of qualitative risk assessment already being ethically difficult supporting this valuation.

Controversy II: Algorithmic bias (2020-2021)

During the fall of 2019, as RISK went public with preliminary results from their first testing of the statistical model in the child protection departments of two municipalities, journalists and data scientists started scrutinising RISK and its algorithm. In January 2020, the algorithm was called out as a "shadow version of the Gladsaxe model" in a tech-magazine (Andersen, 2020), and in social-media platforms, it was criticised in harsh terms for the mere idea of developing algorithms in the field of child protection. In a comments section, it was for instance called a case of "contempt for professional knowledge, incapsulated in tech-utopianism' and 'an assault on the population". When a magazine through the help of a science student found biases in the algorithm, the controversy shifted towards a more technical debate.

The article describes the risk of two kinds of biases. One has to do with the reproduction of biases from the data input – here 'age bias' – which makes the algorithm assume that "the severity of a neglect increases with the age of the child" (Kulager, 2021).[4] The other bias, 'automation bias', has to do with the impact of using the algorithm as decision-support, when or if the caseworkers (uncritically) adjust their decisions to match the risk scores of the algorithm. In the corresponding media debates scientists, professors, social worker, and lay persons dispute the technicalities of the algorithm, the datasets, the validity of the research, the caseworkers' decision-making among several other aspects that touches upon the ethicality of even testing algorithms in the field of child protection. One professor simply judged RISK to be "irresponsible" (Andersen, 2021b). These problematisations thus disputed both the value of the algorithmic model but also the very idea that a decision-support tool could add value to child protection.

The members of RISK reacted to the ethical controversy by participating in the media debates – explaining, arguing and providing answers about how and what they have done, and which iterations they were considering for the second version of their model. They emphasised their willingness to create "openness and transparency about such a difficult subject as the use of machine learning in social work." (Andersen, 2021b). Meanwhile, in June 2020, they marked their publicly available reports with a stamp saying: "Temporary brief. Expired June 2020" and wrote new reports for the next phase of the project. Whereas the expired versions of reports are written in Danish, the new versions are framed in an open, English and theoretically grounded language, cementing their scientific ambitions.

During this revision, the dual valuation of helping children and researching the value of algorithmic decision-support shifted towards the latter. As the project manager explained: "Our hope is that we will be used as a knowledge base, somewhere in the debates, and not be put in a corner as those proposing that we should assess notifications by running them through an algorithm" (interview, project manager, March 2022). This valuation, on the one hand, enacts the algorithm as a valuable object of scientific scrutiny and, on the other hand, it positions RISK as ethically more legitimate than the other ongoing algorithmic experiments since it is the only one being done as research. As a project

member noted during a conversation: "algorithms are too dangerous to simply let loose in practice, without research about the consequences" (field notes, June 2022). Thus, although the algorithm might be dangerous when tested on real cases, the valuation here emphasises the importance of researching algorithms before they are 'let loose'.

Controversy III: Legal uncertainty and lack of legitimacy 2021-2022

During summer of 2021, a scholar of social- and administrative law laid out her take on the legality of employing machine learning algorithms to support child protection services in the media (Andersen, 2021a). Scrutinising the legality of RISK's aim to test the algorithm on real cases, she disputed their legality report, written by the state attorney who had approved the algorithm as legal. A main objection concerns the legal requirement to individually consider which information about a child to collect from a principle of data minimisation; a requirement rendered impossible by the standardised algorithmic collection and assessment of various data points.

In view of this critique, the research group reached out to the state attorney and asked them to re-assess their own audit. In March 2022, the state attorney reaffirmed their earlier assessment, i.e., that they had legal backing for testing the algorithm on real cases. However, rather than finding solace in the authority of the state attorney, the project members at this point began to emphasise a different aspect of this report, leading them to doubt the legality of testing the algorithm on real cases. The project manager explains:

> Our judgement is that it [the state attorney's approval] is simply not solid enough ground for the project to stand on. (…) They [the state attorney] leave several doors open – for instance 'under the conditions of agreement in the field' – and we don't see that [agreement]. Our judgement is, firstly, that we will not continue with the project if there is any doubt about the legality. (…) Another position would be to say: 'well, if the state attorney says it is legal, then there is no doubt about the legality'. But then we know, we will be the object of even more criticism than we have already been, right? We do not wish to be in that position again. (Interview, project manager, June 2022)

As the quote also illustrates, the ongoing critique of their research project points to the lack of legitimacy and due to this, it is not clear cut whether there is a legal basis for testing the decision-support model on real cases. In this regard, the legal rupture in 2021-2022 affords a change in the tectonics of the ethical plateaus of AI for child protection where disagreement about the legal basis becomes the starting point for new interpretations of what is possible to do with algorithms, even if this happens under the label of research.

Even though RISK ends up not testing the algorithm in practice, due to its doubtful legality, the algorithm itself continues to live in a new version where it will only be tested in what the project manager calls "safe environments" – i.e., with artificial data in experimental workshops with child protection caseworkers. This cements a valuation where the algorithm is purely a research object and is delegated the role of acting as guinea pig in a laboratory like setting. Without access to 'real life' cases, it will have no influence on children's lives. The (human) project members, in turn, are delegated the role as researchers, constructing and controlling the artificial setting in which the algorithmic model is to be examined. This marks a shift in the ethical plateaus where algorithms are considered too dangerous to be used for decision-support.

Municipality X: The algorithmic detection of acute notifications

In 2019, two other municipalities began algorithmic developments, in the wake of the Gladsaxe controversy. Even though they were not subject to public controversy, the very existence of these, as we will see, shaped their ideas about what was ethical to do with algorithms. Here, we focus on the algorithmic development in municipality X.

The purpose of the third algorithmic model was to screen emails with notifications of concern for children's wellbeing and identify notifications needing acute responses – the so-called "red notifications" (project description) and in this way prioritise their assessments of notifications. Recalling their initial idea, the project manager

explained that they chose the most vulnerable citizen group (children) for algorithmic experimentation because "this is where we have the greatest potential of being able to help (…) and because, there is a big volume of data with about 7000 notifications per year. This [assessment of notifications] is, of course, a large and difficult task, every day" (Interview, project manager, January 2022). Articulating its value, the Head of Strategy and Governance characterised the algorithmic model as a "super smart person [Kloge-Åge], like a senior employee, who has accumulated knowledge. (…) An artificial person, the caseworkers can consult, as a support" (Interview, Head of Strategy and Governance, January 2022).

From the onset, the project was very aware of the risk of public controversy. As the evaluation report stated:

> Together with the public affairs department, the project has developed a so-called preparedness [beredskab], to answer the many questions we anticipated to emerge. At this point in time, the media had been circulating stories about the use of AI in relation to case work, e.g., from Gladsaxe municipality. The project was therefore very attentive to having a model that supported [email] categorisation, and in no way supported the execution of interventions. (Internal document)

The valuation of this algorithm is at once formulated positively, as an administrative help in handling the large amount of notifications, yet also negatively in terms of what it was *not*, i.e., the Gladsaxe model.

This practice of 'un-ethicising', i.e., positioning themselves as 'ethical' in comparison with others 'unethical' practices (cf. Tønnesen, 2009), continued when RISK became subject to the bias controversy in 2020. Interviewed by the tech magazine Version2, a critical voice in both Gladsaxe's and RISK's public controversies, the Head of Strategy and Governance stated: "We are more oriented towards cleaning for bias than I think they were in RISK. And our project does not make use of the profiling of citizens or predictions [compared to RISK]" (Internal document).

RISK's controversy in 2020 also produced a concern about the risk of automation bias, characterised by the evaluation report as a situation where "human judgements unconsciously lean toward the categorisations generated by the AI, and thus creates an unintentional effect" (internal document). This led them to ensure that the "AI-models' categorisations to have the least possible impact on the caseworkers' decisions" (Interview, Head of Strategy and Governance, October 2022). Thus, instead of visualising the acute-labelled notifications during decision-making, they ran the algorithm as a so-called 'shadow process' and showed the caseworkers the algorithmic classifications during weekly meetings, *after* notifications had been prioritised. In this regard the algorithm was less valued for its ability to provide support during decisions and more as an opportunity to learn and reflect about what is possible to do with algorithms (Interview, project manager and Head of Strategy and Governance, January 2022). This valuation also hinges on it being a 'non-decision model' as the usage of algorithms for decision-support was deemed unethical.

This ambivalent valuation is reflected throughout the evaluation report, our interviews with the project manager and Head of Strategy and Governance. Rather than defining the algorithmic model as something specific, both the project manager and the head of strategy and governance in the municipality emphasise that their main goal is to learn: "what can we do with AI?". The Head of Strategy and Governance continuous:

> This is the cool thing about the project. We get to investigate what is possible and where are the boundaries? (…). The purpose was to (…) feel the boundaries of what is applicable, what is acceptable, what is meaningful and so on. (interview, Head of Strategy and Governance, January, 2022).

This process of adjusting according to "what is acceptable" demonstrates the shifting ethical plateaus to not only be a theoretical concern but a very practical one. It establishes the algorithm as a valuable means to take part in the drawing the boundaries of this new field of innovation in public administration.

With the final evaluation report concluding a rather low accuracy as 0well as a lack of trust

from case workers, it was decided to keep the model running in shadow mode, which, apart from enabling organisational learning, had the assets of "retaining knowledge and competences in the IT-department, which, as a consequence, will be better equipped for working with AI, also in other units" as well as "retaining the possibility of eventually further developing the AI model [for future implementation]" (internal document). Thus, in the end, its valuation has been solidly reconfigured as a (vague) future potential, both in terms of the model itself but also in terms of the municipality's AI competencies.

Municipality Y: Algorithmic sorting of emails – the new colleague

Also in 2019, another municipality (Y) prepared for an experiment in the field of child protection. This municipality also used machine learning to develop an algorithm to detect notifications in an email inbox and to analyse their acuteness. In a meeting agenda in 2020, the algorithmic model was valued for its capacity to "[save] employees the time used to search for notifications [in the mailbox]" (internal document). Indeed, the algorithm was described in meeting agendas and power point presentations as a "mail sorting programme". In an interview with the project manager in charge of the project, we were told that they specifically wanted to "avoid the mistakes that others [Gladsaxe and RISK] have made" (interview, Project manager, November 2021) when deciding how to design their model. She explained these mistakes as 1) the merging of data from different welfare departments (Gladsaxe) and 2) using the predictive capacity of algorithms (Gladsaxe and RISK). "Therefore", she added, "[our algorithm] only collects data, which relates directly to the function it has" (interview, Project manager, November 2021), i.e., searching and marking emails with notifications containing words indicating acuteness. And she relates their choice of model to municipality X as she underscores: "We do not attempt to prioritise". Thus, rather than expanding human analytical capacity, the algorithmic model was trained to do the same as the caseworkers, only faster. It is the speed and not the scope which is articulated as valuable. This minimalised model showcases how the Gladsaxe controversy established clear boundaries of what *not* to do. Correspondingly, we here see how the ethical plateaus of AI for child protection shift towards benefitting administrative work rather than the child and its family.

Despite this starting point, the ethicality of the mail-sorting programme was a concern to begin with. They were particularly concerned with the risk of the algorithm making mistakes in its mail sorting. As they wrote in a meeting agenda: "Whereas "ordinary" IT-systems can "take care of themselves", machine learning demands more ongoing maintenance" (meeting agenda, October 2020, citation marks in the original). Correspondingly, they enrolled a team of skilled, administrative caseworkers to test the algorithmic model in what they termed a "hyper care period" of three months with careful attention to its precision. For this purpose, they personified the algorithm as a new colleague, naming it 'Naomi Notifications'. Below, the project manager describes how she introduced this 'new colleague' to the test team:

> As someone who, well, doesn't care if she sleeps at night and who doesn't go to the toilet, doesn't need food, and that sort of thing. (…). We always say that when we [introduce the algorithm] ...But, because it is also...if you have a challenge with a turnaround in employees, then you can say: "we have a technology who absorbs data in the same way as an employee". It [the algorithm] is a way of consolidating knowledge (…). We try to explain to them that it [the algorithm] can become a very experienced employee who remembers well and can work fast. But to begin with, it isn't. It is more like having an intern. (Interview, project manager, November 2021)

Besides of evoking the algorithm as a person in need of care and training, the personification also enacted the algorithm as a future potential rather than a problem-solving tool for the present. In this valuation, the figure of a new, untrained colleague mobilizes the algorithm as unfinished and full of beginners' mistakes. And, as in the case with interns, the future potential is only achieved through the professional involvement of the human employees, critically scrutinising the algorithm's work.

An evaluation report, however, concludes that this strategy failed, because the enrolled case-workers, instead of providing the anticipated feedback on its precision, they were simply "viewing it [the algorithmic model] as a 'search function' – as in a word document – and correspondingly, they think it cannot be better than it already is" (internal document). The evaluation report concludes: "there are no signs of impact on practice in any particular way". Nonetheless, the algorithm continues to run, and the project manager explains to us that she views it more as a curiosity project, which helps them learn, both how the algorithm and the employees react when machine learning algorithms are employed in practice. This shifts the valuation of the algorithm from being a potential time-saving administrative tool to one that can teach them about employees' use of algorithmic models.

Coinciding with this, the European Court of Justice gives their verdict on the so-called Schrems II[5]-case, establishing the use of American cloud services as non-GDPR compliant. Running on Microsoft cloud services, the municipality therefore kept the algorithm from being implemented in other departments. In spring 2022, a new manager took over in the social service department and started to enquire about the costs of running the algorithm. In an email to us by the project manager, she narrated her estimation that 40.000 kroners per year for a tool that does not make a difference is a lot of money (email, June 2022). They decided to stop the algorithm entirely. As a side comment, in an interview the project manager mentioned that the team had stopped using the algorithm before it was paused because it had begun marking the emails incorrectly (interview, project manager, June 2022). Indeed, the algorithm suffered from the lack of care and trainings. Yet, rather than reconfiguring its potential value, the combination of GDPR-compliance and its lack of positive impact resulted in it being devalued as a mere expenditure not even worthy of ethical concerns. Thus, while informed by the controversies of the other two experiments, the termination, and de-valuation, of the algorithmic model was the end result of many different agencies: lack of training, poor evaluation and GDPR-compliance.

Concluding discussion

In this paper, we have described the emergence and ruptures of ethical plateaus in Danish child protection services. Analysing the relationship between valuations of algorithmic models as (ethically) good, public controversy, and, eventually, the processes that lead to their termination or revaluation, we gained insight in changing boundaries of what is ethically possible to do with algorithmic models in Danish child protection services. In doing so the paper contributes to calls for "situation-sensitive approaches" in the study of AI ethics (Hagendorff, 2020: 14) and it contributes to STS discussions about the making and configuration of (AI) ethics (Seaver, 2021; Ziewitz, 2019). Below, we discuss two implications of these findings, in terms of (1) the role of public controversy in configuring ethical plateaus and algorithmic development and (2) the relationship between ethical plateaus and the distribution of agency across algorithm and humans.

Firstly, we learn how national media scrutiny is important in contesting the ethicality of child protection algorithms, mobilising the responsible organisations to publicly account for – and hence enact – their algorithmic model as ethically good. Figure 2 provides an overview of the development in valuations of the four models, including public controversies and other events that led to their re-valuation or termination. Here, we see how the first controversy, focusing on the Gladsaxe model, resulted both in the termination of this project but also led RISK to produce documents about the ethicality of their research project. The Gladsaxe controversy also influenced the very formulation of the algorithmic projects in municipality X and Y, in terms of becoming an example of what was *not* ethically acceptable to do with algorithms. These two projects, thus, from the outset limited the scope of their algorithmic models, valuing them in relation to administrative and time-consuming tasks rather than vulnerable children. Similarly, we see how controversy II and III, problematising RISK's plan to test their algorithmic model on real cases, led RISK to re-value the algorithm to the extent that it would no longer be tested on real cases. The controversy regarding algorithmic bias, here automation bias, further led municipality X to keep the algorithmic model running in a shadow

mode, decoupled from case workers' prioritisation practices.

Moreover, municipalities X and Y, not terminated by public controversy, are important indications of what is ethically acceptable at the time of writing. Although developed in parallel in two different municipalities, with similar justifications for initiating their projects and shaped by ideas of what *not* to, their valuations of the algorithmic models ended up quite differently. As it seems, the boundaries of what is possible to do in the ethical plateaus of AI for child protection are narrowed down to a point where the employment of algorithmic models seems to be a future ambition rather than a problem-solving tool for the present. In this regard, the practitioners involved in the two projects were reacting to an ethical obligation to prepare the public administration for an imagined future with AI. Whereas municipality X's model became re-valued as a tool for organisational learning and the retainment of AI competences,

Table 1. Ethical plateaus of AI in Danish child protection services 2017-2023.

	Valuations			
Time	**Gladsaxe model**	**RISK**	**Municipality X**	**Municipality Y**
2017	Valued as helping vulnerable children through earlier interventions	Valued as helping vulnerable children through pre-emption of their maltreatment		
2018	**Controversy I: Algorithmic surveillance**			
		Valued as equally ethical as human decisions to test on real cases		
2019	Controversy 1 cancels algorithmic development		Valued for its potential to save time by prioritising notifications	Valued for its potential to save time by identifying notifications
2020	**Controversy II: Algorithmic bias and risky research**			
		Re-valued for its potential to generate knowledge, hence legitimising testing on real cases	Algorithm reconfigured as shadow process. Re-valued as a learning tool for reflecting about how case workers prioritise notifications.	
2021	**Controversy III: Illegal to test algorithm in practice**			
				Evaluation report: Algorithm is not trained and does not benefit practice
				Re-valued as an opportunity to learn how case workers interact with AI
2022		Re-valued as an algorithm only fit for testing on artificial cases	Evaluation report: Low accuracy. Case workers do not trust algorithm	New manager
			Continues as shadow process. Re-valued as a medium for retaining AI competencies in the municipalities	Terminated. Devalued as unnecessary expenditure
2023				

municipality Y's model was devalued as a useless "search function" and unnecessary expenditure.

What are the implications of these developments for our understanding of controversies role in shaping ethical AI? Following Marres' (2021) call to study public controversy, we agree that public controversies should not simply be analysed as instances of democratic interrogations of technological innovation but also as trial grounds for those developing algorithms. Arguing that public controversies increasingly are becoming strategic resources, "[providing] opportunities for the configuration of new markets", Marres warns against "romantic misunderstandings of how scandals happen"(Marres, 2021: 2). In our study, shaping the innovators' sense of what was ethically acceptable to do with algorithms in child protection, public controversy indeed did much more than render visible relations between science, technology, and society. They inadvertently became a platform for scoping what was possible and acceptable to propose to do with algorithms, the two municipal algorithmic models being rather explicit about this.

At the same time, there was no evidence of public controversy being "purposefully used to organise publics for 'innovation'" (Marres, 2021: 13). On the contrary, our interviews with those subject to public controversy indicate that they experienced it negatively. Indeed, municipality Y was concerned that their very use of the term 'algorithm' would associate them with the other contested algorithmic models, and the members of RISK were concerned of more public critique if they followed the state attorney's assessment of their legality. Thus, public controversy was central in shaping ethical plateaus and fed into processes of adjusting the algorithms to what seemed acceptable, but they never became strategic platforms for the innovators.

The second implication of our analysis regards the powers and agency distributed to algorithms. During the emergence of the ethical plateaus, where algorithms were primarily articulated as a solution, most agency was delegated to them. The Gladsaxe model was granted the proactive role of detecting children at risk through prediction, thus intervening *before* any concern or symptom had been registered. Indeed, in encountering critiques of surveillance during rupture 1, the municipality argued for the ethicality of the algorithm by highlighting its 'black hole' processing of data as more ethical than human processing. Compared to the contemporary focus on transparency and responsibility, this articulation may seem absurd but is indicative of the optimism characterising the emergence of the ethical plateaus in Denmark. RISK's decision-support model delegated less agency to the algorithmic model, articulating it as a support *in* human decision making after humans had detected children at risk. Yet, with more ruptures and critique, their enactments of the algorithmic model changed over time, firstly, emphasising the necessity to *research* algorithmic models in use rather than simply implement them without proper evaluation, and later, they decided not to test their model on real cases, removing any risk of harming real casework. Finally, the two municipalities, which initiated their projects after the first public controversy, decided from the onset to limit the agency of their algorithmic models, keeping them from profiling and predicting citizens and from interfering in decision-processes. Whereas one municipality decided to limit the agency even more by reconfiguring it as a possible background algorithm, which they could consult for purposes of reflection *after* decision-making, the second municipality entirely closed their model.

Thus, we see how changing ethical plateaus and a growing awareness of potential controversiality has the effect that gradually less agency is delegated to the algorithmic models, both across the cases but also within those that have been reconfigured. This speaks to Lee and Helgesson's (2020) observation that valuations of algorithmic processes are entwined with distributions of agency across human and system. Thus, ethical plateaus not only shape what is ethically possibly to do with algorithmic models, they also influence on how much agency is given to algorithms – both with regards to how proactive they are *vis-à-vis* humans but also in terms of the roles they are envisioned to have.

Compared to the established literature on AI ethics, with its focus on (developing and assessing) ethical principles for AI, our approach allowed us to explore the relationship between

public controversy's contestation of the ethicality of algorithms and enactments of algorithmic models as good or valuable. This analytical move changes ethics from principles, against which organisational practices can be evaluated, to situated valuation practices. Valuations do not come before (as regulation or involvement) or after (as evaluation), they are part of the very development of algorithms – in this regard the valuation shapes the algorithm to an extent where it can also kill the algorithm by devaluing it as unethical, illegal or as a mere expenditure. This requires us to approach 'the ethical' as a process that is incomplete, uncertain and situated. While our movement towards situated ethics obviously could be criticised for deflating the concept of ethics and for destabilising important efforts to scrutinise algorithms critically and holding them accountable, we propose the contrary: Namely, the endeavour to understand ethics as an emic concept, that is, how organisations mobilise, negotiate and enact the ethically good algorithm. This can teach us important lessons about the lived organisational realities of algorithmic ethics and may potentially make it easier to understand failed attempts at implementing ethical principles and how the figuration of what is ethical is central in the co-constitution of AI and society.

Acknowledgements

Funding from the VELUX FOUNDATIONS (Algorithms, Data & Democracy Jubilee Grant) supported the writing of this article.

References

Andersen T (2020) Gladsaxe-modellen spøger: Nyt AI-projekt skal forudsige mistrivsel hos børn. 29 January. Version2. Available at: https://www.version2.dk/artikel/gladsaxe-modellen-spoeger-nyt-ai-projekt-skal-forudsige-mistrivsel-hos-boern-1089919 (accessed 3 March 2021).

Andersen T (2021a) Jurister: Børneopsporing med kunstig intelligens er ikke forenelig med lovgiving. Available at: https://www.version2.dk/artikel/jurister-boerneopsporing-med-kunstig-intelligens-er-ikke-forenelig-med-lovgiving (accessed 15 January 2023).

Andersen T (2021b) Professor advarer mod kommunal udrulning af børne-AI: »Jeg synes, det er uforsvarligt« | Version2. Available at: https://www.version2.dk/artikel/professor-advarer-mod-kommunal-udrulning-af-boerne-ai-jeg-synes-det-er-uforsvarligt (accessed 15 January 2023).

Andrejevic M (2017) Digital Citizenship and Surveillance| To Pre-Empt A Thief. *International Journal of Communication* 11: 18.

Beer D (2009) Power through the algorithm? Participatory web cultures and the technological unconscious. *New Media & Society* 11(6): 985–1002. DOI: 10.1177/1461444809336551.

Callon M (1986) Some elements of a sociology of translation: domestication of the scallops and the fishermen of St Brieuc Bay. In: Law J (ed) *Power, Action and Belief: A New Sociology of Knowledge?* London: Routledge, pp. 196–223.

Cuccaro-Alamin S, Foust R, Vaithianathan R and Putnam-Hornstein E (2017) Risk assessment and decision making in child protective services: Predictive risk modeling in context. *Children and Youth Services Review* 79: 291–298. DOI: 10.1016/j.childyouth.2017.06.027.

Danish Government (2018) Et Danmark uden parallelsamfund - ingen ghettoer i 2030. Available at: https://www.regeringen.dk/nyheder/2018/ghettoudspil/ (accessed 31 May 2022).

Dencik L, Redden J, Hintz A and Warne H (2019) The 'golden view': data-driven governance in the scoring society. *Internet Policy Review* 8(2). Available at: https://policyreview.info/articles/analysis/golden-view-data-driven-governance-scoring-society (accessed 4 May 2021).

Douglas-Jones R (2017) Making Room for Ethics. *Science & Technology Studies* 30(3): 13–34.

Eubanks V (2018) *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York, NY: St. Martin's Press.

Fortun K and Fortun M (2005) Scientific Imaginaries and Ethical Plateaus in Contemporary U.S. Toxicology. *American Anthropologist* 107(1): 43–54.

Hagendorff T (2020) The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines* 30(1): 99–120. DOI: 10.1007/s11023-020-09517-8.

Helgesson C-F and Muniesa F (2013) For What It's Worth: An Introduction to Valuation Studies. *Valuation Studies* 1(1): 1–10. DOI: 10.3384/vs.2001-5992.13111.

Jørgensen RF (2023) Data and rights in the digital welfare state: the case of Denmark. *Information, Communication & Society* 26(1): 123–138. DOI: 10.1080/1369118X.2021.1934069.

Kjær J (2018) Regeringen vil overvåge alle landets børnefamilier og uddele point, *Politiken* March 3rd. Available at: https://politiken.dk/indland/art6365403/Regeringen-vil-overv%C3%A5ge-alle-landets-b%C3%B8rnefamilier-og-uddele-point (accessed August 11 2023).

Kristensen K (2022) Hvorfor Gladsaxemodellen fejlede - om anvendelse af algoritmer på socialt udsatte børn. *Samfundslederskab i Skandinavien* 37(1): 27–49. DOI: 10.22439/sis.v37i1.6542.

Kulager F (2021) Kan algoritmer se ind i et barns fremtid? I Hjørring og Silkeborg eksperimenterede man på udsatte børn. Available at: https://www.zetland.dk/historie/s8YxAamr-aOZj67pz-e30df (accessed 15 January 2023).

Lee F and Helgesson C-F (2020) Styles of Valuation: Algorithms and Agency in High-throughput Bioscience. *Science, Technology, & Human Values* 45(4): 659–685. DOI: 10.1177/0162243919866898.

Marres N (2021) No issues without media, The changing politics of public controversy in digital societies. In: Swati J and Wasko J (eds) *Media: A Transdisciplinary Inquiry*. Chicago: Intellekt Books, pp. 228-243.

Meilvang ML and Dahler AM (2022) Decision support and algorithmic support: the construction of algorithms and professional discretion in social work. *European Journal of Social Work* 0(0)1–13. DOI: 10.1080/13691457.2022.2063806.

Mittelstadt BD, Allo P, Taddeo M, Wachter S and Floridi L (2016) The ethics of algorithms: Mapping the debate. *Big Data & Society* 3(2): 1–21. DOI: 10.1177/2053951716679679.

Puig de la Bellacasa M (2010) Ethical doings in naturecultures. *Ethics, Place & Environment* 13(2): 151–169. DOI: 10.1080/13668791003778834.

Ratner HF and Elmholdt KT (2023) Algorithmic constructions of risk: Anticipating uncertain futures in child protection services. *Big Data & Society* 10(2): 1–12. DOI: 10.1177/20539517231186120.

Redden J (2018) Democratic governance in an age of datafication: Lessons from mapping government discourses and practices. *Big Data & Society* 5(2).

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

Russell J (2015) Predictive analytics and child protection: constraints and opportunities. *Child Abuse & Neglect* 46: 182–189. DOI: 10.1016/j.chiabu.2015.05.022.

Seaver N (2021) Care and Scale: Decorrelative Ethics in Algorithmic Recommendation. *Cultural Anthropology* 36(3): 509–537. DOI: 10.14506/ca36.3.11.

Statistics Denmark (2022) www.statistikbanken.dk/UND1, (accessed November 11th 2022).

Tønnesen C (2009) *Ethicising: Towards an Understanding of Ethics as Material Practice.* PhD Thesis. Oxford: University of Oxford.

United Nations Department of Social and Economic Affairs (2022) UN e-Government Surveys. Available at: https://publicadministration.un.org/en/Research/UN-e-Government-Surveys (accessed 9 August 2023).

Villumsen AM and Søbjerg LM (2023) Informal pathways as a response to limitations in formal categorization of notifications in child and family welfare. *Nordic Social Work Research* 13(2): 176–187. DOI: 10.1080/2156857X.2020.1795705.

VIVE (2022) Fakta om anbringelser. Available at: https://www.vive.dk/da/temaer/anbragte-boern-i-danmark/fakta-om-anbringelser/ (accessed 12 December 2022).

Ziewitz M (2019) Rethinking gaming: The ethical work of optimization in web search engines. *Social Studies of Science* 49(5): 707–731. DOI: 10.1177/0306312719865607.

Notes

1 Recognizing that the term AI has many competing definitions, we here use it to signify algorithmic systems developed with the use of machine learning techniques.

2 Of course, compared to our brief history of changing ethical boundaries, geological plateaus develop over longer periods and may thus appear more stable.

3 We use the term "internal document" for references that we have as part of our research data, but which we cannot make public due to issues of confidentiality.

4 One explanation for the age bias is that schools send more notifications than daycare institutions. This means that the data set of notifications have an overweight of children going to school. This, of course, does not have anything to do with the situation of the child.

5 The ruling implicated that EU customers of US cloud services, such as Microsoft, Amazon and Google, were themselves responsible for assessing the risks of data being accessed by third party countries as well as verifying the data protection laws of the recipient country – a task impossible to achieve for a Danish Municipality (note from the Data Protection Officer of the municipality, 2021).