IADIS International Journal on WWW/Internet
Vol. 22, No. 2, pp. 47-60
ISSN: 1645-7641

"OF COURSE AI DISCRIMINATES": IDENTIFYING COMMUNICATION GAPS AND CROSS-DISCIPLINARY TRANSLATION CHALLENGES

Hilde G. Corneliussen[1], Gilda Seddighi[2] and Cheshta Arora[1]
[1]*Western Norway Research Institute. Box 163, 6851 Sogndal, Norway*
[2]*Norwegian Research Centre (NORCE). Box 22 Nygårdstangen, 5838 Bergen, Norway*

ABSTRACT

The paper argues for the need to foreground anti-discrimination as a distinct lens when assessing the social impacts of AI design, development and deployment in real-life situations. The argument is based on a survey of around 200 organisations in the public sector in Norway as well as 19 in-depth interviews and presents the challenges of translating 'discrimination' as a socially relevant concept across disciplines and discursive contexts. The paper presents six discursive responses to the risk of discrimination in our study to foreground how focusing on discrimination allows one to address unique challenges that other concepts such as bias and privacy cannot address. By distinguishing concerns around discrimination from other ancillary concerns, such as bias and privacy, we present the need to ground our critical understanding of AI design, development and deployment in actual practices and situations and urge AI developers to actively adopt an anti-discriminatory lens in their practices without replacing it with ancillary concepts such as bias, differentiation, privacy, or other mainstream concepts such as justice or ethics.

KEYWORDS

Artificial Intelligence, Public Sector, Norway, Risk of Discrimination, Inter-Disciplinary Challenges

1. INTRODUCTION

The public sector in highly digitalized societies, such as Norway, collects a lot of information about the citizens every day: information that grants access to services, benefits, and payments from the government to each individual in the country. Who is entitled to social security? Sick pay? Student loans? These and many other questions are assessed based on personal data that the government holds about the citizens. How well is the public sector equipped to mitigate risks of discrimination when individuals' information becomes data for artificial intelligence?

There's an increasing interest among public administrators in Norway and the EU to explore the potentials of data-driven technologies such as artificial intelligence (AI) for a more efficient public sector, more value creation in the business sector and a simpler everyday life for most people (Kommunal- og moderniseringsdepartementet (KMD, 2020). While AI can bring notable benefits to the operations of various actors in the public sector (Alhosani & Alhashmi, 2024; Chiariello, 2021; Wirtz et al., 2019), it also brings risks of producing unfair results and discrimination against different demographic groups (Jørgensen, 2023; Kuziemski & Misuraca, 2020). In the interviews with employees involved in AI development in the public sector in Norway, our initial question about the risks of discrimination according to the Equality and Anti-discrimination Act in Norway was translated into various other concepts, such as bias, fairness, explainability, openness, transparency, ethics, and privacy, or, in some cases, it was not on the agenda. In this paper, we foreground that although there is increasing awareness of various types of negative consequences and risks of AI, one of the barriers to counteract such risks is the challenge of translating concepts and communicating across disciplines and discursive contexts. This challenge became apparent in our study of plans for introducing AI in the public sector in Norway and the risks of discrimination therein (Corneliussen et al., 2022).

Using the interviews from this previous study with public sector employees engaged in the development of AI as our starting point, we will discuss how the concept of discrimination was translated and communicated and what effects and consequences this might have for the use of AI in the public sector. The research question pursued here is: How is the notion of discrimination perceived and translated in the public sector and what are the consequences of prevailing ways of dealing with this risk?

The paper foregrounds that it is important to develop our understanding of how the notion of discrimination is received in the field of AI development and innovation in the public sector to better identify strategies for counteracting the harmful effects of AI in real-life situations. Here, by AI we mean a range of technologies that can be used to automate processes and decision-making with an acute awareness that challenges concerning discrimination will vary depending on the type of AI technology used, the sector in which it is deployed as well as the nature and size of an organization. For instance, data-driven machine-learning technologies present new challenges vis-à-vis discrimination even when used with seemingly harmless, non-personal data, which can still be used to have a negative social impact (Hagendorff, 2019). Similarly, the nature and size of an organization will also influence the design and deployment of the AI system. For instance, a smaller organization at the municipality level will not always have the resources to develop AI systems from scratch and will most likely opt for off-the-shelf AI solutions or third-party AI-as-service solutions (Corneliussen et al., 2022). This dependency on third-party solutions will raise different concerns vis-à-vis discrimination and the necessary social or legal norms for its prevention as opposed to a larger organization that has the resources to design and develop AI solutions from scratch. Thus, it is important to foreground discrimination as a concern that is different from other ancillary concerns within the field of AI such as privacy, fairness and transparency and should be at the forefront of studies assessing AI design and deployment in real-life situations and practices.

The Norwegian government's endeavours concerning data-driven AI are still in its nascent stage (Broomfield & Reutter, 2021) and the journey from ideation to implementation is best described as a "hop-on, hop-off" ride with several challenges producing unwanted stops and exit points (Corneliussen et al., 2024, p. 1). Previous research has examined different aspects underpinning this gap between ideation and implementation and its several consequences while noting the need for more fieldwork-based studies that can identify the "issues that are largely

unseen by both policymakers and practitioners" (Broomfield & Reutter, 2021, p. 73). The list of societal harms and risks associated with AI is continuously growing (*AIAAIC - AI Algorithmic Risks Harms Taxonomy*, n.d.), highlighting an urgent need to reimagine ethical, participatory designs and actively interrogate normative values and practices that inform the development of data-driven technologies. Several examples and case studies have noted how AI models mirror existing discriminatory patterns in society with the risk of further amplifying and accelerating them (Birhane, 2021). Among the many infamous examples so far include deploying AI to identify a new employee or a new top leader for a company, both of which are prone to favour the demographic group stereotypically considered to be more suitable over other demographic groups (Drage & Mackereth, 2022; Srinivasan & Chander, 2021).

The concern with AI becomes more acute when it is discussed in the context of the public sector which relies on mining personal data to deliver essential welfare services. Norway has a long history of registering data about citizens, which can provide a valuable dataset for developing AI-based services (Broomfield and Rutter, 2021). However, the public sector's access to data has made researchers ask whether Norway risks "Stumbling into an Algorithmic Welfare Dystopia" in which predictive models can increase the risk of discrimination against certain demographic groups (Broomfield & Lintvedt, 2022, p. 1). While the vast amount of data that the Norwegian state holds about citizens could provide important input for developing AI, this also introduces risks and is challenged by the privacy provisions under GDPR and its principle of data minimization which states that as little data as possible is stored for as short a time as possible (Malek, 2021). Thus, uncritical use of AI in the public sector can deny essential services to vulnerable and marginalized groups, cause discrimination at scale and give legitimacy to societal issues such as racism, discrimination, and inequality (Keyes et al., 2021). It is within this context that the translation of the concept of discrimination, as outlined in legislation, into concepts and practices that are meaningful to public sector employees becomes particularly important and requires critical unpacking.

Below we will present the empirical data from the study of AI in the Public sector in Norway and the methodological and theoretical framework for this paper before presenting the findings regarding various translations of the concept of discrimination. Thereafter, we will discuss the connotations of these translations for AI development in the public sector before concluding with a discussion on the effects and consequences of the translation.

We foreground the implications that the translations across disciplines and discursive contexts have on the discourse of inclusion and discrimination in the context of AI development in the public sector in Norway. We understand translation not just as a search for equivalence or loss of 'correct' meaning but translations as sites of judgement and continuous contestation (Law & Lin, 2017) that generate meaning (Sarukkai, 2013), actions, and weave partial connections (Strathern, 2004). In doing so, we shift the focus from the individual beliefs and value systems of interviewees. We aim to highlight that the design and development of emerging technologies, such as AI systems and their socio-technical imaginaries, involves a dynamic interplay of forces and circumstances of different situations and disciplines. Within this interdisciplinary framework, actors may translate concepts suddenly or without careful consideration. However, an awareness of these processes can direct us towards more 'careful' (de la Bellacasa, 2011) practices concerning the development and deployment of AI in the public sector.

2. METHODS

The empirical data we use for this paper originates from a research project commissioned by the Norwegian Directorate for Children, Youth and Family Affairs to provide a better knowledge base for efforts to prevent discriminatory effects when using artificial intelligence in the public sector in Norway (Corneliussen et al., 2022). To learn more about how AI was being used and understood as well as plans for developing AI projects in the public sector in Norway, we invited, together with other partners (ibid.), nearly 500 public sector organisations to respond to a survey. These were state and municipal enterprises from sectors such as healthcare, education, employment and welfare administration, tax, customs and police who were invited to respond to the survey which covered questions related to their use and plans for using AI and risks of discrimination. 200 of these organisations responded to the survey.

The study was particularly focused on how public sector organisations perceived and dealt with risks of discrimination when developing AI and therefore a total of 19 in-depth interviews were subsequently conducted with organisations that had an AI system or project that involved the use of personal data which intensifies the risk of discrimination against certain demographic groups. It was during the in-depth interviews, where the question concerning discrimination was discussed with the interviewees, that the key moment of translation appeared – when discrimination was translated and understood by the respondents as a question concerning bias, privacy or other ancillary concepts.

3. THEORETICAL FRAMEWORK

While the original study (Corneliussen et al., 2022) aimed to explore the extent of public sector organisations' use or plan to use AI, this paper takes as its starting point the translations experienced during the interviews when the interviewer asked questions about discrimination. In the study, it was explicitly stated that we used the concept of discrimination in accordance with the Norwegian Equality and Anti-discrimination Act (EAD). EAD accounts for unlawful discrimination of certain demographic groups, with the particular awareness of discrimination as a result of gender, pregnancy, care responsibility, ethnicity, religion, disability, sexuality, gender identity and age. Despite this clear frame of reference for the concept, in most of the 19 interviews, the interviewers observed instances where the concept of discrimination was substituted with some of the other concepts that we will outline below.

To make sense of this moment, where the initial concept was moved into a different frame of reference, we engage with the concept of discursive resources (Dick, 2004; Corneliussen & Seddighi, 2020), which points to how organisations and individuals embedded therein use their own discursive context to understand and make sense of, for instance, external requests to follow rules regarding gender equality. This means that instead of identifying how certain translations are misplaced within the framework of the EAD, we aim to widen our understanding of how such translations involve a different set of discourses from which the interviewees speak. The analysis below thus aims to deepen our understanding of how AI developers in the public sector perceive the risk of discrimination in different ways, which will help identify strategies to counteract this risk in the development of AI in the public as well as private sectors.

While the framework of discursive resources supports the analysis of how interviewees address discrimination, emphasizing the contextual, and discipline-based relevance of certain concepts over 'discrimination', a feminist perspective supports examining how discourses also involve elements of hierarchical relations and power struggles (Livholts & Tamboukou, 2015). The concepts that we choose to use when discussing for instance the risk of discrimination, will affect not only how this risk is perceived but also the ability to deal with such risks. These two theoretical perspectives will guide our analysis: first, in understanding the discursive context of the translations encountered in the interviews, and second, in pointing towards the effects and consequences of such translations for the development of AI for the public sector.

4. FINDINGS

In the survey as well as the interviews, the issue of discrimination was presented within a framework of the Norwegian Equality and Anti-Discrimination Act (EAD) in the following manner: "According to the Norwegian Equality and Anti-Discrimination Act, it is not permitted to discriminate based on the categories listed below. Which of these are relevant to or handled by AI in your organisation? Gender, pregnancy, leave in connection with childbirth or adoption, caregiving responsibilities, ethnicity, religion/belief, disability, sexual orientation, gender identity, gender expression, age". By explicitly associating our understanding of discrimination with EAD, we wanted to emphasize that we were interested in the risk of unlawful discrimination which could lead to directly or indirectly treating certain demographic groups differently or unfairly compared to others. However, even though the starting point within the framework of the EAD was made clear from the start, the interviewees often responded directly through other concepts, such as bias or privacy, as the basis for understanding the question. Furthermore, most of the interviews were made in Norwegian, making this moment of translation even more peculiar. The Norwegian equivalent for 'discrimination' is 'diskriminering' but the English term 'bias' has no Norwegian equivalent apart from the imported term 'bias', and some of the concepts introduced rather reflected a disciplinary terminology from computer science or data science. Below we have grouped the responses under six headings that reflect slightly different contexts and backgrounds, or discursive resources.

Discrimination: Only in a couple of interviews did the respondents speak about discrimination with reference to its definition as 'unlawful discrimination' following the EAD act. These respondents were part of AI projects that had interdisciplinary expertise on discrimination either within the group or at the organisation level. Other respondents frequently invoked discrimination through other ancillary concepts such as:

Bias: Most respondents evoked 'bias' in response to the question concerning discrimination. This mainly reflected a computer science or a data scientist's epistemic universe where the concept of bias is used to talk about how algorithms, AI models, and data can produce unfair results (Srinivasan & Chander, 2021).

Mainstream concepts: The second most common way of responding to questions about discrimination was to respond within a reference frame of one or more of a series of concepts that often appear in discussions of harmful results of AI, thus what we have labelled as mainstream concepts here. These were concepts such as fairness, openness, transparency, explainability, representativity, justice, and ethics. These concepts were used to talk about the

challenges of avoiding harmful results with the use of AI and machine learning in particular and reflect concepts dominating these debates today (Birhane et al., 2022).

Privacy: The European Union's General Data Protection Law, GDPR, received a lot of attention when it was introduced in 2016 and replaced the EU's 1995 directive on privacy Data Protection Directive. Among other things, for the public sector, GDPR draws up a hard legal boundary concerning the collecting, storing and processing of personal data within the EU. In the interviews, this was frequently reflected in concerns around dealing with personal data where the main focus was on compliance and the need to operate within the limits of this law. The law itself poses some challenges to AI development such as requesting to collect as little data as possible and storing it for as short as possible. However, in some cases, GDPR appeared to take the place of related laws such as the EAD and replaced concerns around discrimination with privacy.

Differentiation: The concept of differentiation was used interchangeably with the concept of discrimination by some of the interviewees. In one of the interviews, the interviewee responded to the question about discrimination by saying "Of course we discriminate. We need to differentiate between different people such as men and women" (paraphrased). This informant spoke within the framework of mathematics, statistics, and computer science, where it is necessary to differentiate between different entities (Friedman, 1997).

Unaddressed: In a couple of our interviews, it was made clear that the issue of unlawful discrimination against certain demographic groups was not on the agenda of the project. One of the informants simply replied by saying, "We haven't thought about that. Thank you for reminding us". They illustrate a group of AI developers who had not had access to competence about discrimination in the development process, and therefore they had not come to think of this as an important perspective to include in the process.

5. DISCUSSION: EFFECTS AND CONSEQUENCES

The discursive universe encountered when asking AI developers in the public sector about the risk of discrimination with AI reflects how this question has a variable status in the different AI projects. In this section, we discuss the consequences of each of the discursive responses:

Bias: Given that computer scientists were prominent in the interviews, bias was the language most familiar to them when questions of discrimination were raised. Bias and discrimination are, however, not synonyms and the jump from bias to discrimination is not straightforward. The concept of 'bias' in the Anglo-phone philosophy of science has witnessed a series of reformulations. It is no longer sufficient to merely understand bias as preferences, deviations, inclinations or prejudices that can be accounted for to arrive at a value-free or neutral science. Bias has a moral import which can have discriminatory effects. Feminist approaches to techno-science have played a leading role in interrogating how different types of bias underpinning knowledge-making practices might have discriminatory effects (Longino & Doell, 1983). This helped debunk the common myths of science such as objectivity, universality and value-neutrality by foregrounding that technoscience is often situated and partial. Thus, it is important to account for the discriminatory effects when accounting for bias in science.

More recently, the question of bias has emerged as central to interrogating the theory and practice of data science and consequently how the discourse around AI performs the same modern myths of objectivity and value-neutrality that were earlier identified in the practice of

science. These mythologies of scientific practice that data science mimics co-produce bias as a problem that can be treated through methodological fixes. The perspective of a technological fix to bias in society can make it challenging to address the discriminatory outcomes of AI. This is because it creates the illusion that simple technological solutions can make the problem disappear. The foundation for bias in sight might be limited to easily recognized variables such as gender, age, and sexuality while failing to explore and deal with the fact that discrimination is dynamic and all attempts at AI-related monitoring, predicting, automating, and profiling can lead to several discriminatory outcomes. This was apparent in some of the methods used to 'wash' the data, such as anonymizing it and removing basic features like gender and age. However, it remains an open question to consider whether other pieces of data could also potentially lead to discriminatory AI results. Furthermore, the concept of bias was most often used to refer to unconscious bias, that is, discrimination that we are not aware of (Suveren, 2022). In the interviews this appeared as a translation from the risk of discrimination with AI to a general risk of discrimination in society, often combined with the questioning of whether AI can be better than society. The findings from the interviews were supported by the results of the survey, where less than a third of the survey respondents reported that the main challenge of dealing with discrimination is demographic bias in the data, while two-thirds of respondents answered that the main challenge was to anonymize the data (ref (Corneliussen et al., 2022), Figure 3.1).

Mainstream concepts: The mainstream concepts and values such as fairness, openness, transparency, explainability, representativity, justice, and ethics have a critical role in the development of AI. However, they do not cover the responsibilities covered under EAD. Furthermore, it seems that these concepts, due to their prominent role in the discourses and policies about AI, are not only concepts that AI developers in Norway are more familiar with than EAD's version of discrimination but also appear as more tangible and accessible tools for dealing with harmful effects of AI. In the interviews, this was illustrated with these concepts not only replacing the concept of discrimination but also representing a solution. However, these mainstream concepts do not always align with the aims of the EAD or capture the dynamic, intersectional aspects of discrimination in real-life situations. Concerns around discrimination are not always already implied in these mainstream concepts, and AI developers relying on these mainstream concepts still need to develop a careful understanding of the risk of discrimination underpinning their AI projects.

Privacy: As part of the EEA agreement, GDPR was introduced in Norway in 2018 as an unyielding framework with mechanisms for ensuring that companies fulfil their obligations according to the law. In our study, 58% of the respondents believed they had a high level of competence in data protection (Corneliussen et al., 2022). However, as mentioned previously, our questions about discrimination were often interpreted as challenges related to privacy and the processing of personal data as laid out in GDPR, rather than the EAD Act. This confluence of challenges related to discrimination with privacy and consequently GDPR is not unique to our respondents but can be observed more broadly in the larger discourse on AI accountability and decision-making which has historically been dominated by privacy scholars (Gillis & Simons, 2019). Easy access to information was one of the core promises of the Internet. However, it also served to further discriminatory practices and within this context, privacy, or control over one's personal information became increasingly relevant for preventing discrimination (Roberts, 2014). This symbiotic relationship between privacy and discrimination requires a nuanced approach in the context of algorithmic decision-making, especially when concerns around privacy are replaced with GDPR which is first and foremost, a legal tool to

govern the data protection practices of companies and is often experienced as a bureaucratic tool to comply with (Simonsen & Fürst, 2024) rather than to think broader implications of AI, privacy and algorithms discriminating effects.

However, it should be stated that in its formulation, GDPR, explicitly foregrounds the link between privacy, data protection and discrimination, unlike the 1995 Data Protection Directive which did not show any awareness of the relationship between privacy and discrimination (Calvi, 2023). Thus, in complying with GDPR, AI developers must consider the negative risks of profiling on the grounds of "racial or ethnic origin, political opinion, religion, or beliefs, trade union membership, genetic or health status or sexual orientation" (Calvi, 2023). In this regard, GDPR owes its popularity to being the "first legislation to address discrimination explicitly" (Goodman, 2016) which also explains the confluence between GDPR and discrimination that we encountered in our study.

More recently, however, legal and conceptual analyses of the GDPR have pointed out several limits and effectiveness in combating algorithmic discrimination (Calvi, 2023; Gillis & Simons, 2019; Goodman, 2016). First, algorithmic discrimination is not defined under GDPR and it doesn't adequately differentiate between disparate treatment (i.e., discrimination on grounds of special categories) and disparate impact (i.e., discrimination occurring despite neutral practices) (Goodman, 2016). While the language of *Recital 71,* where negative profiling is prohibited, appears to inhibit disparate impact, mechanisms within GDPR such as data sanitization (removal of special categories before processing) and algorithmic transparency or the right to explanation, work towards eliminating disparate treatment (Goodman, 2016). This inconsistency and skewed focus on disparate treatment, thus, limit GDPR's effectiveness in tackling algorithmic discrimination head-on, as it becomes a tool to comply with without considering the risks of the disparate impact of algorithmic practices. Second, in light of ongoing breakthroughs in algorithmic practices and data-based technologies, a limited focus on privacy and disparate treatment inhibits our ability to fully grasp the risks of discrimination, especially concerning those approaches that do not rely on personal data but can extract sensitive or discriminatory information from seemingly unsuspicious streams of data (Hagendorff, 2019).

Third, GDPR and other legal tools operate with an individualist, control-based notion of privacy which doesn't capture the interdependent experience of privacy concerning today's digital technologies where your friends' consent to share their information is taken as consent enough to process your information in their contact list (Hagendorff, 2019). Within this world of digital interdependence, privacy is at best a collective effort and a narrow focus on privacy is doomed to fail. Finally, and most importantly for our discussion, neither does GDPR explicitly and unambiguously cover other grounds for discrimination such as gender or age nor does it account for an intersectional nature of discrimination which is a glaring lack (Calvi, 2023). To account for intersectional discrimination is to account for the unique experience of discrimination that some groups face which cannot be sorted under separate categories. This means that while automated systems may appear fair with respect to sensitive attributes considered separately and would seemingly comply with GDPR mandates, they might perpetrate discrimination against groups situated on unfavourable intersections (Calvi, 2023).¨

In the context of Norway and contrary to GDPR, EAD has long been in the making for more than four decades in Norway and was informed by feminist approaches to the transformative potential of law. EAD was premised on the assumption that the "obligation of public authorities to prevent discrimination and promote equality under the EAD Act will ensure that gender equality is mainstreamed into all laws, policies and practices" (Hellum et al., 2024). In 2018, several laws were merged into one general equality and anti-discrimination act with a stated

objective "of improving the position of women and minorities" (Hellum et al., 2024, p. 145) and dismantling disabling barriers created by society. Overall, despite some discrepancies, the act pivots upon a broad and intersectional notion of discrimination (Hellum et al., 2024) which can make it a good starting point to tackle a range of discriminatory cases in AI. However, while GDPR is legally binding, the EAD, although recognised and accepted in society in general, does not have a mechanism for ensuring that organisations comply with the law or follow-up guidance for organisations in need (Nordberg, 2019). Given our discussion, there's an urgent need to address the limits within GDPR and further probe how other relevant laws and mandates such as EAD can help address the discriminatory risks of algorithmic decision-making.

Differentiation: The translation of discrimination into differentiation also changes the meaning from the EAD targeting unlawful discrimination, to a lawful differential treatment. This translation thus makes it very difficult to talk about discrimination according to the EAD. Interpreting discrimination as differential treatment shifts the conversation away from negative effects to the need to represent real-world scenarios as they are.

Unaddressed: It should be sufficient to foreground that not putting questions relating to discrimination on the agenda does not promote or facilitate any strategies or ways of dealing with issues of discrimination.

Table 1 sums up the different translations, their discursive context and meaning, and the effects and consequences they might have.

Table 1. Translations of the concept of discrimination referring to the Equality and Anti-Discrimination Act

Translation of discrimination	Discursive context and meaning	Effects and consequences
Discrimination	Equality and Anti-Discrimination Act	Ability to deal with unwanted discrimination.
Bias	Commonly used in computer science for referring to differences between demographic groups in society	Dealing with demographic differences through a technological fix. Risk of simplifying both problem and solution
Mainstream Concepts	Common concepts to talk about challenges of producing fair results with the use of AI.	Alternative frameworks talk about harm and justice, without engaging with the dynamic and emergent nature of discrimination.
Privacy	Legal framework for working with personal data	Taking precedence before other considerations such as EAD.
Differentiation	A mathematical concept for distinguishing between variables.	Replacing discrimination for differentiation erases the core principle of unwanted discrimination targeted by the EAD.
Unaddressed	The issue of discrimination was not on the agenda for the AI project.	Not having discrimination on the agenda makes it difficult to deal with such issues.

6. CONCLUSION

The paper presents our analysis of communication challenges when interviewing AI developers in the public sector in Norway. When asked about their views on the risk of discrimination, most informants exchanged this concept for other concepts. As we have illustrated above, it is possible to understand this translation in terms of the background and contexts that the informants were talking from. However, these translations also have effects and consequences.

The biggest challenge is that perspectives that can help nuance our understanding of unlawful discrimination were weakly represented among AI developers in the public sector in Norway. Not including this perspective will make it more difficult to fully understand the risk of discrimination, and thus also to develop strategies to avoid further producing or reproducing discrimination. While the mainstream concepts, as well as GDPR, are important tools for AI development, an uncritical recourse to them will not fully capture the ambition of the EAD, and thus they cannot fully substitute the necessary critical work required in any development and deployment of AI projects. Unfortunately, it seems the mainstream concepts in some cases appear as a fully adequate solution to any challenges of reducing harmful results with AI. Thus, while we appreciate these mainstream concepts' frameworks as highly relevant for AI development, we will encourage AI developers to also include a critical perspective on discrimination along with the understanding that the EAD is premised upon, which is discrimination against certain demographic groups for reasons that are unlawful and thus not legitimate.

This study and the analysis above illustrate that we lack a common language to communicate across disciplines and fields of expertise. One example is the exchange of discrimination with the concept of differentiation, moving away from unlawful discrimination to deliberately identifying differences between individuals.

This furthermore illustrates how knowledge about discrimination in terms of EAD is necessary to be involved in AI development. In other words, AI development needs to happen in interdisciplinary groups that can critically unpack the existing as well as emergent risks of discrimination associated with AI.

Bias is clearly the most used concept that AI developers talk about to avoid the harmful effects of AI models. As we have argued above, this will, however, not fully cover the need to focus on unlawful and emergent discrimination in accordance with the Equality and Anti-Discrimination Act. It has been argued that most data-based AI models are error-prone (Pasquinelli & Joler, 2021) and harms (such as discrimination, and exclusion) can be introduced throughout the ML life (Suresh & Guttag, 2021) in the way data is collected, curated, and annotated, or how the AI system is designed, by whom, pivoting on which values, and to achieve what goals. Given this, an evocation of 'bias' plays a performative role in the practice of data science. Critical perspectives on data-driven approaches have underscored how reliance on bias reinforces the notion of neutrality and objectivity in data science (Birhane et al., 2022). This reliance reproduces a false sense of security, or strengthens the hope that we can develop technologies for good if we can fix its inherent bias or align it with abstract principles such as transparency or fairness (Powell et al., 2022). Further, it shifts the focus away from systemic harms of technologies we build to individual perception. Within critical work on AI and data-driven practices, several authors have argued for a move away from shallow, technical, methodological considerations of bias to its social, and ethical considerations where bias ought to be understood in relation to AI harms such as discrimination, exclusion and inequality

(Birhane, 2021; Dancy & Saucier, 2022; Draude et al., 2018, 2019). Thus, nurturing an attitude of responsibility concerning bias in AI needs a sociotechnical approach that addresses the cultures of algorithm (Draude et al., 2019), and a stronger, interdisciplinary and critical focus on how these systems are designed, developed, and deployed (Dancy & Saucier, 2022). It also demands deep insights into individual and organisational behaviour, economic incentives, as well as complex dynamics of the sociotechnical systems (Adomavicius & Yang, 2022). This requires improving disciplinary skills or fostering interdisciplinarity within teams working on data-driven solutions in fields as varied as hate speech, agriculture or climate (Doman & Garrison, 2021). These interventions underscore responsibility and response-ability vis-a-vis bias rather than ways to eliminate it (Feinberg, 2007).

Similarly, given the political history of digital technologies, privacy and control over one's information have become a central site of struggle. However, with ongoing breakthroughs in digital technologies and the overreliance on an individualist notion of privacy, a narrow focus on ensuring privacy, via the mandates of GDPR, might deter developers from addressing concerns related to discrimination in and beyond AI. While there is a symbiotic relationship between privacy and discrimination, one does not ensure the other. There is a need to think of discrimination in addition to privacy-respecting design when assessing the design, development and deployment of AI in different sectors and organizations.

In Norway, discrimination is addressed through the Equality and Anti-discrimination Act which has been in the making for more than four decades and has directed social change via the "transformative potential of law" (Hellum et al., 2024, p. 135). While more critical legal research is needed to understand if EAD is enough to tackle AI or data-based discrimination, it's important to note that the act's emphasis on discrimination and on avoiding new and emergent discrimination gets obfuscated in the practice of developing AI for the public sector.

The findings of our study illustrate that discrimination is not on the agenda of AI developers in Norway and is often side-stepped by evoking other ancillary concerns such as difference, bias, privacy, transparency etc. While it's important to ensure that AI models are fair, transparent and privacy-respecting, encoding these values doesn't guarantee that the model will not have discriminating effects.

ACKNOWLEDGEMENT

The research project referred to in this paper was funded by the Directorate for Children, Youth and Family Affairs. An earlier version of this paper was presented at the IADIS International Conference ICT, Society and Human Beings 2024 (part of MCCSIS 2024) and published in the proceedings of the conference. We want to thank our colleagues Aisha Iqbal and Rudolf Andersen at Rambøll Management Consulting who also participated in the data collection, for collaborating on the project. The elaborate discussion of the translation challenges is original for this paper produced by the authors.

REFERENCES

Adomavicius, G. and Yang, M., 2022. Integrating Behavioral, Economic, and Technical Insights to Understand and Address Algorithmic Bias: A Human-Centric Perspective. *ACM Transactions on Management Information Systems*, Vol. 13, No. 3, pp. 1-27. https://doi.org/10.1145/3519420

AIAAIC - AI algorithmic risks harms taxonomy., n.d. [dataset]. Available at: https://www.aiaaic.org/projects/ai-algorithmic-risks-harms-taxonomy (Accessed: 30 October 2023)

Alhosani, K. and Alhashmi, S. M., 2024. Opportunities, challenges, and benefits of AI innovation in government services: A review. *Discover Artificial Intelligence*, Vol. 4, No. 1, pp. 18. https://doi.org/10.1007/s44163-024-00111-w

Birhane, A., 2021. Algorithmic injustice: A relational ethics approach. *Patterns*, Vol. 2, No. 2, 100205. https://doi.org/10.1016/j.patter.2021.100205

Birhane, A. et al., 2022. The Forgotten Margins of AI Ethics. *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. Seoul, Republic of Korea, pp. 948–958. https://doi.org/10.1145/3531146.3533157

Broomfield, H., and Lintvedt, M.N., 2022. Is Norway Stumbling into an Algorithmic Welfare Dystopia? Snubler Norge inn i en algoritmisk velferdsdystopi? *Tidsskrift for velferdsforskning*, Vol. 25, No. 3. pp 1-15. https://doi.org/10.18261/tfv.25.3.2

Broomfield, H., and Reutter, L., 2021. Towards a Data-Driven Public Administration: An Empirical Analysis of Nascent Phase Implementation. *Scandinavian Journal of Public Administration*, Vol. 25, No. 2, pp. 73-97.

Calvi, A., 2023. Exploring the Synergies between Non-Discrimination and Data Protection: What Role for EU Data Protection Law to Address Intersectional Discrimination? *European Journal of Law and Technology*, Vol. 14, No. 2, pp. 331-334. https://doi.org/10.1007/s10676-019-09510-5

Chiariello, A. M., 2021. European Review of Digital Administration & Law | AI and Public Services: A Challenging Relationship Between Benefits, Risks and Compliance with Unavoidable Principles. *European Review of Digital Administration & Law - Erdal*, Vol. 2, No. 2, pp. 185-203. https://doi.org/9791259947529 16

Corneliussen, H. G. et al., 2022. Bruk av kunstig intelligens i offentlig sektor og risiko for diskriminering. Vestlandsforsking rapport 7–2022, pp. 80. Available at: https://www.vestforsk.no/sites/default/files/2023-03/VFrapport7_2022_KI_i_offentlig_sektor.pdf

Corneliussen, H. G. et al., 2024. Artificial Intelligence in the Public Sector in Norway: AI Development as a Hop-on-Hop-off Journey. *AI, Data, and Digitalization. SAIDD 2023. Communications in Computer and Information Science*, Vol. 1810. Springer, Cham., pp. 160-172 https://doi.org/10.1007/978-3-031-53770-7_11

de la Bellacasa, M. P., 2011. Matters of care in technoscience: Assembling neglected things. *Social Studies of Science*, Vol. 41, No. 1, pp. 85-106. https://doi.org/10.1177/030631271038

Doman, M. and Garrison, C., 2021. Introducing algorithmic bias considerations in an introductory CS course. *Journal of Computing Sciences in Colleges*, Vol. 37, No. 5, pp. 31-42.

Drage, E., and Mackereth, K., 2022. Does AI Debias Recruitment? Race, Gender, and AI's "Eradication of Difference". *Philosophy & Technology*, Vol. 35, No. 4, pp. 89. https://doi.org/10.1007/s13347-022-00543-1

Draude, C., Klumbyte, G., Lücking, P. and Treusch, P., 2019. Situated algorithms: A sociotechnical systemic approach to bias. *Online Information Review*, Vol. 44, No. 2, pp. 325-342. https://doi.org/10.1108/OIR-10-2018-0332

Draude, C., Klumbytė, G., & Treusch, P., 2018. Re-Considering Bias: What Could Bringing Gender Studies and Computing Together Teach Us About Bias in Information Systems ? *CEUR Workshop Proceedings*, Vol. 1-2103, Paper No. 3. Available at: https://ceur-ws.org/Vol-2103/paper_3.pdf

Feinberg, M., 2007. Hidden Bias to Responsible Bias: An Approach to Information Systems Based on Haraway's Situated Knowledges. *Information Research: Proceedings of the Sixth International Conference on Conceptions of Library and Information Science—"Featuring the Future"*, Vol. 12, No. 4. Available at: https://informationr.net/ir/12-4/colis07.html

Friedman, J. H., 1997. On Bias, Variance, 0/1—Loss, and the Curse-of-Dimensionality. *Data Mining and Knowledge Discovery*, Vol. 1, pp. 55-77. https://doi.org/10.1023/A:1009778005914

Gillis, T. B., & Simons, J., 2019. Explanation < Justification: GDPR and the Perils of Privacy (SSRN Scholarly Paper 3374668). https://doi.org/10.2139/ssrn.3374668

Goodman, B., 2016. A Step Towards Accountable Algorithms? Algorithmic Discrimination and the European Union General Data Protection. *29th Conference on Neural Information Processing Systems*, Vol. 9, Barcelona, Spain.

Hagendorff, T., 2019. From privacy to anti-discrimination in times of machine learning. *Ethics and Information Technology*, Vol. 21, No. 4, pp. 331-343. https://doi.org/10.1007/s10676-019-09510-5

Hellum, A. et al., (eds.), 2023. *Nordic Equality and Anti-Discrimination Laws in the Throes of Change: Legal developments in Sweden, Finland, Norway, and Iceland*. Routledge, London. https://doi.org/10.4324/9781003172840

Jørgensen, R. F., 2023. Data and rights in the digital welfare state: The case of Denmark. *Information, Communication & Society*, Vol. 26, No. 1, pp. 123-138. https://doi.org/10.1080/1369118X.2021.1934069

Keyes, O., Hitzig, Z., and Blell, M., 2021. Truth from the machine: Artificial intelligence and the materialization of identity. *Interdisciplinary Science Reviews*, Vol. 46, Nos. 1-2, pp. 158-175. https://doi.org/10.1080/03080188.2020.1840224

Kommunal- og moderniseringsdepartementet (KMD), 2020. *Nasjonal strategi for kunstig intelligens*. Kommunal- og moderniseringsdepartementet (KMD). https://www.regjeringen.no/contentassets/1febbbb2c4fd4b7d92c67ddd353b6ae8/no/pdfs/ki-strategi.pdf

Kuziemski, M. and Misuraca, G., 2020. AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings. *Telecommunications Policy*, Vol. 44, No. 6, 101976. https://doi.org/10.1016/j.telpol.2020.101976

Law, J. and Lin, W., 2017. The Stickiness of Knowing: Translation, Postcoloniality, and STS. *East Asian Science, Technology and Society: An International Journal*, Vol. 11, No. 2, pp. 257-269. https://doi.org/10.1215/18752160-3823719

Livholts, M. and Tamboukou, M., 2015. *Discourse and Narrative Methods: Theoretical Departures, Analytical Strategies and Situated Writings*. SAGE Publications, London, California, New Delhi, Singapore.

Longino, H. and Doell, R., 1983. Body, Bias, and Behavior: A Comparative Analysis of Reasoning in Two Areas of Biological Science. *Signs: Journal of Women in Culture and Society*. Vol. 9, No. 2, pp. 206-227 https://doi.org/10.1086/494044

Malek, M. A., 2021. Bigger Is Always Not Better; less Is More, Sometimes: The Concept of Data Minimization in the Context of Big Data. *European Journal of Privacy Law & Technologies (EJPLT)*, Vol. 1, pp. 212.

Nordberg, T. H., 2019. Arbeidsgivers ansvar for likestilling i arbeidslivet. *Tidsskrift for Kjønnsforskning*, Vol. 43, No. 2, pp. 90-107. https://doi.org/10.18261/issn.1891-1781-2019-02-03

Pasquinelli, M. and Joler, V., 2021. The Nooscope manifested: AI as instrument of knowledge extractivism. *AI & Society*, Vol. 36, No. 4, pp. 1263-1280. https://doi.org/10.1007/s00146-020-01097-6

Powell, A. B., Ustek-Spilda, F., Lehuedé, S. and Shklovski, I., 2022. Addressing ethical gaps in 'Technology for Good': Foregrounding care and capabilities. *Big Data & Society*, Vol. 9, No. 2, pp. 1-12. https://doi.org/10.1177/20539517221113774

Roberts, J. L., 2014. Protecting Privacy to Prevent Discrimination. *William & Mary Law Review*, Vol. 56, No.6, pp 2097-2174.

Sarukkai, S., 2013. Translation as Method: Implications for History of Science. In B. Lightman, G. McOuat and L. Steward (eds.) *The Circulation of Knowledge Between Britain, India and China*. Brill, Netherlands, pp. 309-329. https://doi.org/10.1163/9789004251410_014

Simonsen, J. K., and Fürst, E. L., 2024. Experiences of GDPR in Norway: Politics of autonomy and control. *Anthropology Today*, Vol. 40, No. 2, pp. 18-20. https://doi.org/10.1111/1467-8322.12875

Srinivasan, R. and Chander, A., 2021. Biases in AI Systems: A survey for practitioners. *Artificial Intelligence and Machine Learning: Communications of the ACM*, Vol. 64, No. 8, pp. 44-49. https://10.1145/3464903

Strathern, M., 2004. *Partial Connections*. Altamira Press, UK.

Suresh, H. and Guttag, J. V., 2021. A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle. *EAAMO '21: Proceedings of the 1st ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, No. 17, pp. 1-9. https://doi.org/10.1145/3465416.3483305

Suveren, Y., 2022. Unconscious Bias: Definition and Significance. *Psikiyatride Güncel Yaklaşımlar-Current Approaches in Psychiatry*, Vol. 14, No. 3, pp. 414-426. https://doi.org/10.18863/pgy.1026607

Wirtz, B. W., Weyerer, J. C. and Geyer, C., 2019. Artificial Intelligence and the Public Sector—Applications and Challenges. *International Journal of Public Administration*, Vol. 42, No. 7, pp. 596-615. https://doi.org/10.1080/01900692.2018.1498103