Addressing the ethical principles of the Norwegian National Strategy for AI in a kindergarten allocation system

Stian Botnevik Andrey Belinskiy Edis Asotic
Herman Stoud Platou Thomas Søvik Marija Slavkovik

Abstract

The Norwegian National Strategy for Artificial Intelligence (NNSAI) published in 2020 includes seven principles of ethical AI. This paper explores whether those seven principles are stated in a clear enough way and are feasible to be satisfied by a specific AI system. We build an implementation of the Gale-Shapley algorithm to allocate kindergarten places in Bergen, Norway. The presented solution is then evaluated against the ethical principles from the NNSAI. We argue that it is difficult to respect all the ethical principles when implementing a solution to a matching problem.

1 Introduction

The NNSAI, first published on January 14th 2020 [18], aims at providing *a framework for both public and private entities seeking to develop and use artificial intelligence*. It details applicable definitions and topics such as; data, data management, AI-innovation, explainability, trustworthiness, frameworks, infrastructure, research areas, current competencies, responsible use and security. The strategy concludes with a set of ethical principles that need to be addressed when developing or implementing artifacts that use any form of AI.

> We explore to which extend can the ethical principles in the NNSAI be satisfied by an algorithmic matching system for assigning children to kindergartens.

The kindergarten allocation problem, like a lot of matching problems [13], has a socially sensitive nature. Parents have a preference for their children when it comes to kindergartens, and kindergartens have an obligation to prioritize select children to comply with national legislation and local policies. There exists no national standard regarding how the allocation process should take place, causing practises to vary between municipalities. While national legislation is available through sources such as *Lovdata*[1], these do not offer details about allocation practices on the municipal level.

The municipality of Bergen uses a digital system to manage children's applications, but the allocation is done manually[2]. Allocators use a set of criteria[3] based on children's

This paper was presented at the NIK-2020 conference; see http://www.nik.no/.

[1] `https://lovdata.no`

[2] Disclosed in private communication, see GitHub repository under Appendix

[3] `https://www.bergen.kommune.no/styringsdokument/2821046`

attributes to assign priority. A subset of these criteria gives some children special allocation priority. While we know that certain attributes contribute to an increased priority for a child, exactly which are weighted the strongest, and how they are combined is not necessarily disclosed.

On one hand automating the kindergarten allocation problem has the advantage of reducing costs, saving time, and can enforce precision and transparency of the process. On the other hand, algorithms cannot incorporate the same intuition and common sense that trained professionals can. For a successful and efficient automation of matching, automated algorithmic approaches and AI systems have to be adequately transparent and explainable. Such requirements on AI systems are addressed by the recently drafted ethical principles of the Norwegian National Strategy for Artificial Intelligence (NNSAI). In summary, a successful allocation automation will need to fulfill the following criteria: a) The system needs to allocate children to kindergartens automatically, with minimal human intervention; b) The system needs to conduct said allocation and follow the existing national and local policies, which are considered ethical and fair at the time of the implementation; c) The system needs to adhere to the NNSAI ethical principles.

The allocation of kindergarten places is a problem most commonly modeled as a two-sided matching problem. Such problems are derivatives of the original stable marriage problem proposed by David Gale and Lloyd Shapley in 1962. Because this is well accepted and standard approach [2, 1, 3], we implement a kindergarten allocation system using the *Gale-Shapley algorithm* [9]. Our implementation has been built on available information about how kindergarten allocation is currently being performed and respects most of the legislation and policies applicable to the municipality of Bergen.

The goal of our implementation is not to evaluate how good the Gale-Shapley algorithm is for kindergarten allocation. This is a standard well tested algorithm. We build a prototype that is detailed enough to allow us to test whether the NNSAI principles are feasible to satisfy in a matching type problem. Our contribution is an evaluation of the ethical principles of NNSAI with respect to a concrete example of an AI system.

2 Preliminaries

The Gale-Shapley algorithm

The kindergarten allocation problem is an instance of the well known *stable marriage problem* [13]. The stable marriage problem is the problem of finding a *one to one* match between two sets of individuals. A solution to the problem is proposed by David Gale and Lloyd Shapley in their seminal work from 1962 [9] and it has been widely applied in school contexts [2, 1, 3], but has also been used to solve otherwise difficult and morally sensitive matching problems such as kidney exchange [13].

The kindergarten allocation problem we consider can be classified as a *two sided bipartite matching problem* [13]. Bipartite meaning members of two disjoint sets of agents are being matched, and two sided meaning each "side" has total preferences over the members of the other "side".

The Gale-Shapley algorithm works as follows on the original stable marriage problem. We assume there exist a set of men and a set of women that need to be matched. In the first round, each man proposes to his most preferred woman. If a woman receives more than one proposal, she rejects all except the one that she has the highest preference for among the proposers. The formed pairs are now temporary engagements. In the next round all the rejected men propose to the next woman on their preference list. The women reject all but their top preference from the new proposers, this includes the temporary fiancé. New temporary engagements are (possibly) formed. This round is repeated until all women has received a proposal, when this happens the engagements are declared final and the algorithm terminates. All men and women will be matched if the number of men and women are equal. The resulting match will be one that is best for all the men given they are on the proposing side. This does not mean that every man will get their most preferred woman, but they will get the most preferred available woman that they can. This does not exclude that a man can get their least preferred woman. The reason for this is that all women can have man Z as their lowest preference, and Z ends up having to propose to his least preferred woman after being rejected by all the more preferred women. This woman, if not engaged, will have to accept the proposal. This is still $Z's$ best possible match, since all other have rejected him.

The Gale-Shapley algorithm results in a stable match. Stability is defined as the state when there exits no match where any of the agents can swap their partner with someone else to get a better match [9].

We put the children in the match proponents. The Gale-Shapley algorithm favors the proposing side of the two collections, as noted by Maggs and Sitaraman [16]. In a balanced relationship this can be perceived as unfair, but because of the relation between children and kindergartens, we considered it just to favor the children. The preferences of the kindergatens over the children are derived from the municipal admission criteria.

In addition to stability, we require that the match is also *fair*. We consider that children are treated fairly in connection with allocation to kindergartens places if: "All children are treated in the same objective way in respect to kindergartens admission criteria, and only judged based on these criteria when being allocated to kindergarten places".

Municipal Criteria

The municipal criteria that need to be fulfilled in allocation describe what features children should be judged by, namely prioritised, when being allocated to kindergartens. The features are as follows: children with disabilities, children of guardians with severe visual impairment, children under care of the Child Services, children of minority language, children with single guardians, children with siblings in same kindergarten, children of employees, location (living in same district as kindergarten) and age[4].

3 Proposed System

Our system[5] allocates children to kindergartens using the described Gale-Shapley algorithm. The system takes two inputs: A list of children, each with a list of their preferred kindergartens, and a list of kindergartens. The system outputs a list of all the matches between children and kindergartens, and a *waiting list* if there are more children than available places. The process is displayed in Figure 1. For the matching between children and kindergarten places to work, two criteria needs to be fulfilled: a) There needs to be an equal amount of children and kindergarten places; b) All preference orders, of children over kindergartens and kindergartens over children, need to be total.

If there are more children than there are vacant kindergarten places, we create a waiting list with a size equal to the places missing in the kindergartens. This ensures

[4]The features are translated and simplified admission criteria, gathered from the *statutes for Bergen Municipal kindergartens*: `https://www.bergen.kommune.no/styringsdokument/2821046`

[5]The system, its code and figures describing the system are available in the following GitHub repository - `https://github.com/benq66/ai_ethics_2020`

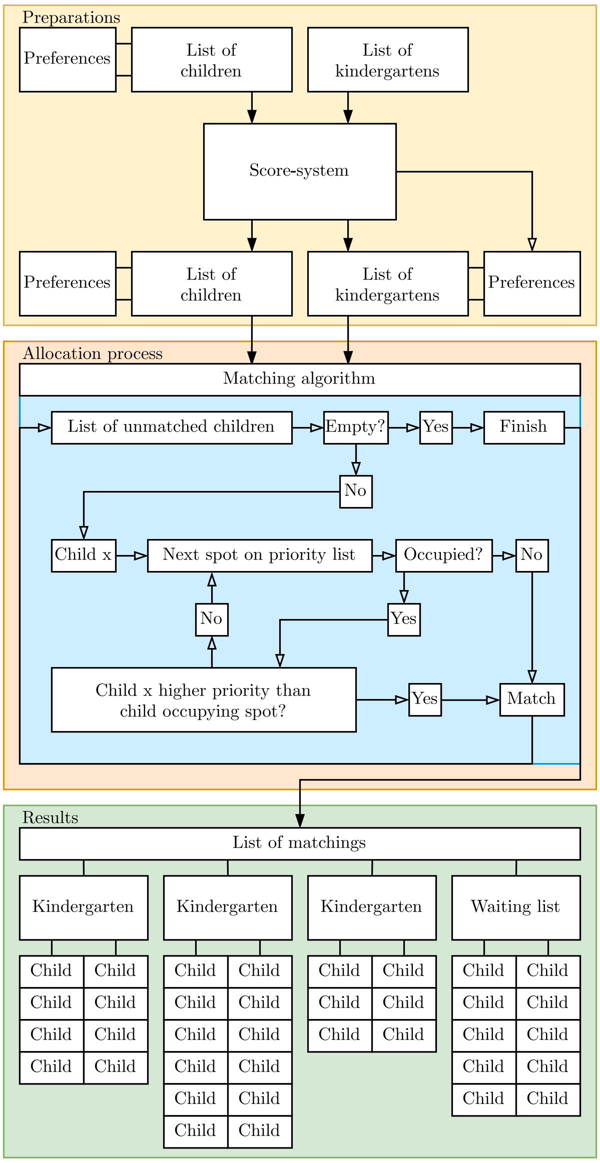

Figure 1: System architecture, with details of the allocation process.

that we have virtually as many positions as we have children, satisfying the first criterion. To satisfy the second criterion, we have created a *scoring system* to build preference lists for the kindergarten places. This is a standard approach because the kindergartens do not have preferences over the applicants in the true sense of the word.

We generate a preference list for each individual kindergarten and then all kindergarten places (in this kindergarten) get a copy of this preference list. This system is built according to the municipality's policy, which is disclosed in the preliminaries section. This system calculates a score for each child by assigning weights to the features described in the policy. The features we use, and the weights we give them are accounted for in the GitHub repository under the *Figures* directory.

The *scoring system* supports further customization by allowing for the assignment of different weights to any number of the children's features. The system calculates a score per child for every kindergarten. The score is used to determine the positions of the children in the preference lists of all the kindergartens. Using this score system, all kindergartens will prefer the children that fulfill the most criteria for their kindergarten.

The kindergarten preference list will in practice be ordered after how many criteria the children fulfill. If two children have the same score then time of application will determine who gets the higher preference. A simplified example of how scoring affects a kindergartens preference list is given in the GitHub repository. The waiting list acts as a kindergarten and its preference list of children is random.

The matching algorithm firstly puts all children in the unmatched list. Then the head of this list is selected, and the availability of its most preferred kindergarten is checked. If the allocation is available then the child is assigned this place temporarily, and the algorithm proceeds to the next entry. Eventually a conflict will arise where the child's preferred spot is occupied by some other child, in such an event their indices in the kindergarten's preference list are compared and the allocation is given to the child with the lowest index, the losing party is then inserted back into the list of unmatched children. Since index position in the preference list is used, there will be no collisions of children with equal score. The algorithm iteratively continues in this fashion until all children have been allocated to a kindergarten place, or assigned to the waiting list. The system traces all steps of the matching algorithm, thus making it possible to investigate why a child was assigned a specific allocation.

Our prototype simulates an allocation of 250 children with random values and priorities. It can be accessed by visiting the following URL – `http://wildboy.uib.no/~kud005/ACID/`. A description of how to use the system is available as well – `https://github.com/benq66/ai_ethics_2020/blob/master/Javascript/guidetoACIDfinal.pdf`.

4 Evaluation

In this section we assess how well our system works in respect to the criteria that the municipality of Bergen is obliged to satisfy. Next, we evaluate our system against the ethical principles of the NNSAI.

System Evaluation

Our implementation of the Gale-Shapley algorithm is child-optimal meaning that the children are given the first mover advantage in the match. To further evaluate the quality of the match we need empirical testing. We can measure the result of the allocation process by examining how many children are allocated to their highest prioritized kindergarten. Our current system, based on synthetic data, assigns a first choice kindergarten to approximately 60 percent of the children. The actual performance of the system would require testing with real data.

It can be difficult to guarantee the interests of the kindergartens when the result of the algorithm is child-optimal. We need to ensure that the kindergartens are not at such a disadvantage that they fail to fulfill their obligations to the national and local governments. This can only be accomplished with a careful crafting of the scoring system that generates the "preferences" of the kindergartens over children. To be readily deployable our system would require expert input and verification of the scoring system in addition to extensive testing over real data.

The Norwegian National Strategy for Artificial Intelligence (NNSAI)

The NNSAI itself focuses on providing an overview of the current state of AI, its components and general application in the context of Norwegian society. The part that we are interested in is included near the end of the document and consists of seven ethical principles regarding the development of trustworthy AI. For the AI-driven system to be considered trustworthy it must be lawful, ethical and robust, according to the European Commission's high-level expert group [7].

We examine each of the seven ethical principles and analyze how applicable they are in regards to the system that we have developed.

1) AI-based solutions must respect human autonomy and control. The first principle consists of two main parts: (i) people should have the right to have an alternative when using the AI-based system that makes impactful decisions on their life if they choose not to use one, and (ii) that people should be included in the process of making said decisions, and provide feedback on all the stages of this process.

The first part of this principle is not related to our system since the availability of an alternative human-driven decision-making procedure depends on the municipality. Currently, our system does not have such functionality. However, the final decision of the system should be first verified by individuals before executing it, therefore, the first principle is partially satisfied.

It can be concluded that this principle is mainly focusing on the requirement to include the presence of human individuals in one or more stages of the algorithm, which impacts the development of the AI-based system. Since this principle is specific enough, it is clear what are the exact requirements, therefore the principle seems quite possible to implement.

2) AI-based systems must be safe and technically robust. The second principle is more abstract, it mainly advises to ensure that the system behaves as predictably and reliably as possible. This is achieved by incorporating well-proven, well-established and trustworthy techniques, concepts and software solutions during the development of the system.

In the case of our system this principle is satisfied since the system uses a well-known algorithm that has proven to provide meaningful results due to its *stable matching* property [1]. However, it is not in any case ideal since it requires many additional tweaks and adjustments due to the specific context in which the system is built.

As a result it can be assumed that this principle is a part of a good approach to developing any system and therefore can be considered reasonable.

3) AI must take privacy and data protection into account. The third principle implies the importance of following the "General Data Protection Regulation" (GDPR) [20]. One of the ground rules imposed by GDPR is that "everything you do in your organization must, by design and by default, consider data protection" [6]. It means that a certain number of rules and terms such as data protection principles, accountability and consent should be followed and respected when collecting and processing personal data.

Our system stores the data using the pseudo-anonymization technique, i.e. the entities in the database are identified by pseudonyms instead of personal information. This is considered as one of the ways to comply with the GDPR's part about secure storage of personal data [15]. It should also be noted that most of the GDPR's data protection principles are out of the scope of the current implementation of our system and therefore cannot be reliably evaluated.

The principle, however, is very specific and should be perceived as mandatory when developing a system that processes personal information in any way, shape or form.

4) AI-Based systems must be transparent. This principle argues that a system must be transparent, explainable and traceable. This means that it should be possible for legal persons or individuals to trace how a decision that affects them was made. Explainability can be done in several ways. Some examples are that a system can show what input data was used and how this data affects the results, how precise the system is, or how changes to some instance in the system affects the result [14]. In our system we have implemented an overview of all the children's input data. However this should only be available to the operators of the system. Present issues with our allocation system is that while it would be possible for a parent to see their own input data, it would not be possible for individuals to gain full insight in all input data for other children due to privacy concerns. Several pieces of sensitive information are used, and it would be against national privacy regulations as well as the GDPR to give insight to others into this data. This means that a individual will not be able to see how the whole decision process was conducted, and therefore the principle cannot be implemented to its full extent in our proposed system.

5) AI systems must facilitate inclusion, diversity and equal treatment Principle five discloses that AI must contribute to inclusion and avoid discrimination. In addition, datasets used for training of AI systems should be avoided if they contain historical bias, are incomplete or incorrect. Our system makes decisions solely based on the Bergen municipality criteria for kindergarten admission. We assume that these criteria are developed to promote equality. Employees that conduct manual allocation of children can be prone to giving preferential treatment to children and families they know, which can lead to unequal treatment. This is assuming that the employees have access to some identifiable data such as names of children/parents. A system such as ours will not be prone to giving such treatment, and therefore conduct a more equal allocation.

6) AI must benefit society and the environment What is beneficial for society is a broad question to answer, as it depends on what is actually considered "beneficial". Automating processes in general can have several implications on a societal level, both positive and negative. Automation can free up the human workforce for a specific task, and these resources can be used on other tasks that are not automated yet, or not possible to automate. These resources can either be laid off, if no alternative work is provided, or they can be used for other tasks that are not possible to automate. A report by Oxford economists states that 47% of total US employment is in the risk of being automated [8]. Automation of tasks does not always lead to negative consequences such as layoffs. As an example, Bergen municipality saves 12 full-time employment positions by automating a simple form filling task, and uses these resources on other tasks to improve services for their citizens [4]. This is an example that shows how automation can be a change in the nature of work. Resources can be allocated to non-automatable tasks which for example requires creativity or social skills [21].

7) Accountability This principle complements the other listed above, and entails the introduction of mechanisms to ensure accountability for solutions built on AI and for their outcomes both before and after the solutions are implemented. This principle does not state who is to be held accountable, but rather states that there must be at least one entity that can be held accountable for eventual wrongdoings by the system. In a system such as ours, several stakeholders are involved, and thus several legal entities can be held accountable such as the municipality, or the creators of the system which could either be "in house" or external. We have not implemented or considered such measures in our system, but we provide an example of a framework below that can be used for implementing such measurements.

Berscheid & Rowewer-Despres [5] propose a three party regulatory framework between *developers* (an entity or person that creates an AI system), *clients* (entity or person that intends to deploy the AI system) and the *subject* (an entity or person that is the target of the deployed system). The client would be the municipality that uses the system, and the subject is the children, or the parents of the children. The framework proposes a regulatory system where a pair of documents needs to be filed for any autonomous decision-making system. The first document consists of an *AI Validation Document (AVID)* which validates the scope of the system and how the systems decisions can be interpreted in the case of an audit. The second document is called *Deployment Disclosure Document (DDD)* and it identifies the process for deploying an AI. This document can help to detect fault if a legal challenge should appear in regards to deployment of the AI. The AVID documents are to be filed first, and be approved to establish the validity of the application. After the AVID has been approved, the DDD document is filed. The benefits of using this framework is that developers share knowledge with their clients, and the clients can have a framework for evaluating and questioning the developers. The framework also allows for open questions towards the institution that deploys the AI, and makes it possible to hold the institution accountable.

5 Related Work

We consider the related work on the topic of fair matching of applicants in the context of kindergartens [12, 22, 10]. Clearly, none of this work considers compliance with NNSAI. There exist other works, which address similar matching problems in the context of higher education, and given this overlap we have chosen to include a selected few of these as well [2, 11, 1]. The works can be categorized according to their scope. While [1, 22, 12] are focusing more on the **implementation** details, [11, 17] are about public **perception**.

Visma [23] is at present developing a solution for automating kindergarten allocation. *Visma Flyt Barnehage* takes an automated approach to kindergarten allocation by implementing the integer programming approach as presented in Geitle et al. [10]. Their algorithm allocates children to kindergartens while respecting children preference and municipal policies. Visma describes current approaches as manual, varying, and extremely time consuming. *Visma Flyt Barnehage (VFB)* offers its solution to a selected few municipalities. The pilot study is being conducted in cooperation with Tønsberg municipality. Geitle et al. [10] was published in early 2020 and aligns greatly with our work. While the subject of Geitle et al. is the same as ours and they use an Integer Programming Method which solves the problem as a constraint satisfaction program, rather than a matching program – they use an objective function to describe the priority of children. They also use three groups of kindergartens, one of them is a dummy kindergarten for those who do not get any place (we do the same). They do not consider NNSAI. Their solution includes a set of additional features, such as gender balance, age balance, travelling time, and partial allocation.

Veski et al. [22] present a redesigned allocation method for kindergartens in Estonia, using the deferred acceptance algorithm. Efficiency is defined as the ability to meet predefined goals of satisfying applicant preferences such as bringing together siblings and minimizing travel. Fairness is based on the idea of equal access invoked from the probability of the child getting their first preference. The algorithm described in the paper creates a child-optimal stable matching, primarily prioritising siblings and distance. The goal of their research was to create a matching policy focusing on efficiency and fairness. Their research resulted in seven different allocation policies.

Abdulkadiroglu et al. and Kennes et al. [1, 12] address existing weaknesses in the systems that allow for *gaming* the matching system by being strategic when assigning priorities for schools. While these systems are implemented in two separate countries, they seemingly have the same fallacy. By changing the child's pre-school the year before they are due to graduate and move up to primary school, one can in effect manipulate primary school admittance. To prevent this the authors present two *strategy proof mechanisms* to replace the existing one: "The deferred acceptance mechanism" (DAM) and "The top trading cycles mechanism" (TTCM). These algorithms were then considered for application in the Boston school system, where in the end the DAM was proven to be favorable.

Jobin et al., and Marcinkowski et al. [11, 17] focus on the perception of algorithmic matching systems and the overarching global landscape of AI ethics guidelines. The discussed matching systems might differ on a local and national level by how much the factors weigh and even by what sort of matching is performed. However, as outlined in [11] there is a global interest in these issues and insurance that they are fair, trustworthy, and transparent. The work presents 11 principles that have been aggregated from a number of publications, with 88% of these publications being from 2017 or later. While the NNSAI is not directly mentioned in this study as this study was published the year before, they are indirectly included. The NNSAI principles are direct derivatives of the principles published by the *EU commissions Independent High-Level Expert Group on Artificial Intelligence*, which are included in the study. The presented principles, listed according to frequency, are: transparency, justice and fairness, non-maleficence, responsibility, privacy, beneficence, freedom and autonomy, trust, dignity, sustainability, and solidarity.

While the NNSAI principles are not directly equivalent to the listed, they can be understood as more generic and therefore encompass the same notions. One worthy mention is that the NNSAI does not explicitly mention beneficence, the principles can therefore be understood as to be harm reducing rather than the opposite. The ethical principles of the NNSAI are[6]: 1. Human agency and oversight, 2. Technical robustness and safety, 3. Privacy and data governance, 4. Transparency, 5. Diversity, non-discrimination and fairness, 6. Societal and environmental well-being, and 7. Accountability.

Marcinkowski et al. [17] present a series of hypotheses that address public perception of automatic decision making (ADM) systems.The findings can be summarized as supporting in the sense that the public has trust in ADM systems, and even favor them over a human decision maker due to inherent bias. The authors conclude that perception of fairness regarding ADM admission systems in universities is very relevant and can significantly impact the reputation of the university, as well as the amount of students that choose to study in this university. This can be extrapolated to our research since the reputation of the communal public services as well as the satisfaction of the parents are at stake depending on the process and the results of the distribution of children between kindergartens.

6 Discussion

System. Our system is at this time not suitable to do all the tasks of kindergarten allocation that is expected by the municipality or law. The system does not satisfy the following two juridical criteria: a) guaranteed allocation for children who turn one year old during the year the child applies for kindergarten; b) guaranteed allocation within their

[6]https://www.regjeringen.no/en/dokumenter/nasjonal-strategi-for-kunstig-intelligens/id2685594/

birthday month for children who turn one year old in September, October or November that year. Additionally, the system does not satisfy the municipal criteria of having an opportunity to appeal the decision of an allocation. Lastly, we have no real world data about the kindergarten children in the municipality, so we are not able to model our mock data to reflect the real child population. However, the system we have is still sufficient to do an initial *evaluation of it using the ethical principles of the NNSAI.*

Ethical implications. Our system does not have any ethical implications that are inherent to the system. As long as the scoring is implemented in a way that reflects the criteria the municipality needs to comply with, there will be no reason for the system to be unfair based on our definition. We therefore argue that the system is fair and further stress the importance of proper justification for the included features and attributed weights. Feature selection would have to be conducted by experts that can provide sufficient justification for the included features and their weights. The system's result will be without negative ethical implications for the children, if the score-system has an ethically sound foundation, which we assume it will have when it is implemented in a real life scenario. The system we have designed is an implicit ethical agent [19], where all ethical considerations are addressed in the system design.

The implementation of a system such as ours could possible cause some employees that work at the municipality to lose their jobs. This is not related to the development of the system or the ethical behavior of the system, but rather a possible consequence of the implementation of the system. In this sense our system is also an ethical-impact agent [19].

By developing our system we allow for the possibility of implementing a far more advanced system of how kindergartens prioritize children, than what is currently used by the municipalities in Norway. It is possible to have an almost unlimited number of factors in the decision process, instead of just the few that are utilized now. The possibility to make a more advanced score-system can cause ethical implications. For example, politicians can enforce policies that affect the score-system, and result in more diversified or segregated kindergartens. This raises the question of who is responsible for this system, the creators or those who manage it, or both. By making this system, we at least act as facilitators of these possibilities.

7 Summary

We here considered an algorithmic matching system for assigning children to kindergartens and explored to which extend it maintains the ethical principles in the NNSAI. We demonstrated that our system is successful at solving the fundamental processes of the kindergarten allocation problem. We also discuss each of the seven ethical principles of the NNSAI. The first principle in the NNSAI is hard to implement in a system that uses a stable matching algorithm. The nature of the algorithm does not allow for people to reserve themselves against being handled by it, as is the desire of the first NNSAI principle. The stable match algorithm needs an equal amount of children and kindergarten places to be able to work, so what happens to the children that does not want to be handled by the algorithm? We cannot handle them manually since this can be unfair to the rest of the children, since they can get a favourable position when not handled by the algorithm. We cannot put them directly in the waiting list (last priority) since this will neglect the value of the principle, and in practice will not give the child (or its parents) a choice of autonomy and control. The NNSAI guidelines accept that not all principles can be implemented in an AI system, and that in this case a good reasoning and documentation should be provided. Our reason for not implementing this principle is that it conflicts with the nature of the algorithm, and the person who wishes not to be handled by the algorithm will be put in a disadvantage.

We can implement a mechanism for the people that do not want to be handled by the algorithm and leave it at that, even if they get a disadvantage, just so that we comply with the ethical principles. But we do not consider this a good solution as it goes against the intent of the principle. This leaves us with the ironic conclusion that it can be more ethical not to follow the ethical principles of the NNSAI. Conflicts do also arise in regards to the fourth ethical principle that addresses the requirement of transparency, traceability and explainability. Allowing any subject insight into how decisions were made in the system by disclosing the input data could violate the national privacy regulations, GDPR, and thus also conflict with the third ethical principle that addresses privacy and data protection. We cannot implement all the principles without exceptions. Following from this, the use of a stable matching algorithm is maybe not the best way to solve the automatic allocation problem in regards to complying with the ethical principles of the NNSAI. We can make compromises and it is up for discussion if our reason for not implementing the first principle is acceptable. If this is acceptable, we can conclude that a stable matching approach works for making a kindergarten allocation system. If not, the NNSAI recommends to discard the system approach.

How applicable are the ethical principles of the NNSAI for creating an ethically compliant AI system? As an advisory framework we consider the principles as partly applicable for creating an ethically compliant system. It is not helpful or clear in all areas or with specific challenges, however. For example, in assisting the development of our score-system the principles are of some help, they advise to facilitate inclusion, diversity and equal treatment. However, they do not state what to weigh the heaviest of these three concepts. Since they are partly contradictory (diversity and equal treatment), it is not very helpful for establishing specific development goals. Either way we are left with the impression that these principles can be somewhat helpful, but can also be used for legitimating ethically questionable AI systems. As we saw earlier in the discussion it was possible to have an (arguably) unethical system, while complying with the principles.

Our research resulted in two contributions, the first is in regards to automating the children allocation process in Norwegian kindergartens, and the second is providing an evaluation of the NNSAI ethical principles. Future work can address the limitations of our proposed system by testing it with real applicant data, requirements and weights. In regards to NNSAI, it would be interesting to investigate an implementation of our system that satisfies all ethical principles. Particularly to find a good way to address the principle of respecting human anatomy and control.

References

[1] Abdulkadiroglu, A., Pathak, P., Roth, A.E., Sonmez, T.: Changing the boston school choice mechanism. Tech. rep., National Bureau of Economic Research (2006)

[2] Abdulkadiroğlu, A., Pathak, P.A., Roth, A.E.: The new york city high school match. American Economic Review **95**(2), 364–367 (2005)

[3] Abdulkadiroğlu, A., Sönmez, T.: School choice: A mechanism design approach. American economic review **93**(3), 729–747 (2003)

[4] Agency, N.D.: Robotisering i offentlig sektor – lær av digifrid! `https://www.difi.no/opplaeringstilbud/nett-tv-og-presentasjoner/3082017-robotisering-i-offentlig-sektor-laer-av-digifrid` (2017), [Online; accessed 07-May-2020]

[5] Berscheid, J., Roewer-Despres, F.: Beyond transparency: a proposed framework for accountability in decision-making ai systems. AI Matters **5**(2), 13–22 (2019)

[6] Commission, E.: What is gdpr, the eu's new data protection law? https://gdpr.eu/what-is-gdpr/ (2018), [Online; accessed 07-May-2020]

[7] Commission, E.: Ethics guidelines for trustworthy ai. https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines (2019), [Online; accessed 06-May-2020]

[8] Frey, C.B., Osborne, M.A.: The future of employment: How susceptible are jobs to computerisation? Technological Forecasting and Social Change **114**, 254–280 (2017). https://doi.org/10.1016/j.techfore.2016.08.019

[9] Gale, D., Shapley, L.S.: College admissions and the stability of marriage. The American Mathematical Monthly **69**(1), 9–15 (1962)

[10] Geitle, A.H., Johnsen, Ø.K., Ruud, H.F.E., Fagerholt, K., Julsvoll, C.A.: Kindergarten allocation in Norway: An integer programming approach. Journal of the Operational Research Society **0** (2020). https://doi.org/10.1080/01605682.2020.1727786, https://doi.org/10.1080/01605682.2020.1727786

[11] Jobin, A., Ienca, M., Vayena, E.: The global landscape of ai ethics guidelines. Nature Machine Intelligence **1**(9), 389–399 (2019)

[12] Kennes, J., Monte, D., Tumennasan, N.: The day care assignment: A dynamic matching problem. American Economic Journal: Microeconomics **6**(4), 362–406 (2014)

[13] Klaus, B., Manlove, D.F., Rossi, F.: Matching under preferences. Handbook of Computational Social Choice pp. 333–355 (2016). https://doi.org/10.1017/CBO9781107446984.015

[14] Liao, Q.V., Gruen, D., Miller, S.: Questioning the ai: Informing design practices for explainable ai user experiences. arXiv preprint arXiv:2001.02478 (2020)

[15] Lindquist, J.: Data science under gdpr with pseudonymization in the data pipeline. https://www.dativa.com/blogs/data-science-gdpr-pseudonymization-data-pipeline/ (2018), [Online; accessed 08-May-2020]

[16] Maggs, B.M., Sitaraman, R.K.: Algorithmic nuggets in content delivery. SIGCOMM Comput. Commun. Rev. **45**(3), 52–66 (Jul 2015). https://doi.org/10.1145/2805789.2805800, https://doi.org/10.1145/2805789.2805800

[17] Marcinkowski, F., Kieslich, K., Starke, C., Lünich, M.: Implications of ai (un-)fairness in higher education admissions: The effects of perceived ai (un-)fairness on exit, voice and organizational reputation. In: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. p. 122–130. FAT* '20, Association for Computing Machinery, New York, NY, USA (2020). https://doi.org/10.1145/3351095.3372867, https://doi.org/10.1145/3351095.3372867

[18] Ministry of Local Government and Modernisation: The national strategy for artificial intelligence. https://www.regjeringen.no/en/dokumenter/nasjonal-strategi-for-kunstig-intelligens/id2685594/?ch=1 (2020), [Online; accessed 05-May-2020]

[19] Moor, J.: The nature, importance, and difficulty of machine ethics. IEEE Intelligent Systems **21**, 18–21 (08 2006). https://doi.org/10.1109/MIS.2006.80

[20] Parlament, E., Council: Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation) (2016), http://data.europa.eu/eli/reg/2016/679/oj

[21] Pham, Q.C., Madhavan, R., Righetti, L., Smart, W., Chatila, R.: The impact of robotics and automation on working conditions and employment [ethical, legal, and societal issues]. IEEE Robotics & Automation Magazine **25**(2), 126–128 (2018)

[22] Veski, A., Biró, P., Põder, K., Lauri, T.: Efficiency and fair access in kindergarten allocation policy design. The Journal of Mechanism and Institution Design **2**(1), 57–104 (2017)

[23] Visma: Visma optimization technologies. https://www.visma.com/artificial-intelligence/optimization/ (2020), [Online; accessed 22-April-2020]