Ethical decision-making in human-automation collaboration: a case study of the nurse rostering problem

Vincent Bebien, Odile Bellenguez, Gilles Coppin, Anna Ma-Wyatt, Rachel Stephens

Ethical decision-making in human-automation collaboration: a case study of the nurse rostering problem

Vincent Bebien[1,2*], Odile Bellenguez[1], Gilles Coppin[3], Anna Ma-Wyatt[2], Rachel Stephens[2]

[1] IMT Atlantique, LS2N, La Chantrerie, 4 rue Alfred Kastler, 44307, Nantes, France.

[2] School of Psychology, University of Adelaide, Adelaide SA 5005, Australia.

[3] IMT Atlantique, Lab-STICC, Technopôle Brest-Iroise CS 83818, 29238, Brest, France.

*Corresponding author(s). E-mail(s): vincent.bebien@imt-atlantique.fr;

Abstract

As artificial intelligence (AI) is increasingly present in different aspects of society and its harmful impacts are more visible, concrete methods to help design ethical AI systems and limit currently encountered risks must be developed. Taking the example of a well-known Operations Research problem, the Nurse Rostering Problem (NRP), this paper presents a way to help close the gap between abstract principles and on-the-ground applications with two different steps. We first propose a normative step that uses dedicated scientific knowledge to provide new rules for an NRP model, with the aim of improving nurses' well-being. However, this step alone may be insufficient to comprehensively deal with all key ethical issues, particularly autonomy and explicability. Therefore, as a complementary second step, we introduce an interactive process that integrates a human decision-maker in the loop and allows practical ethics to be applied. Using input from stakeholders to enrich a mathematical model may help compensate for flaws in automated tools.

Keywords: AI Ethics, Decision-Making, Operations Research, Nurse Rostering Problem

This version of the article has been accepted for publication, after peer review but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: http://dx.doi.org/10.1007/s43681-024-00459-w

1 Introduction

In order to develop AI more thoughtfully, recent approaches have tried to frame what would be safe and ethical uses of AI and what should be avoided. [16, 40, 60] The AI Act, proposed in the European Union, is a recent example of laws aiming to regulate AI development by identifying potential risks in applications [46]. At the same time, institutions and scholars have been researching how to design ethical AI by publishing different strategies, with the goal of defining good practices and guiding AI developers [40].

Among all of the applications encompassed by the term 'AI', Operations Research (OR) algorithms are used in many fields to find optimal or near-optimal solutions, mainly for industrial problems. As opposed to Machine Learning (ML) applications, which can be defined as 'data-driven', OR applications are called 'model-driven'. In the first case, a model is built by a black-box algorithm using a huge amount of data given to it as input. This model could then be used, for instance, to predict values from incomplete incoming data that would best fit the initial data set. In contrast, OR models are designed by developers using mathematical modelling, their own understanding of the problem, and discussions with decision-makers who will use the model afterwards (see [62]). By describing precisely which solutions are feasible, a solver can be used to find one of them, following some criteria. Epistemic concerns as described by Mittlestadt et al. [48] might not be the same for ML and OR applications, as many of them are focused on data quality for the former, whereas the latter are focused on responsibility and decision-making processes [5, 11, 67]. However, they both constitute decision-making tools that may deeply impact the environment for which they make decisions [52]. Thus, the consequences of decision-making algorithms may be analysed in the same manner regardless of the technology they use, making OR applications relevant to AI ethics.

Recently, the field of AI ethics has been criticised as ineffective for many reasons, one of them being its principlist approach to ethics. Rességuier and Rodrigues [57] describe this approach as 'toothless' because it can be easily misused and considered a 'softer version of law' [40]. Munn [49] argues these principles are meaningless, not conveying any concrete statement about what should be done, and isolated, insufficient to address profound issues in the industry. Mittelstadt [47] also states AI ethics has mainly produced 'vague, high-level principles and value statements' and highlights the challenge of translating these principles in practice.

In section 2, we propose a different way to approach ethics that starts with context-field knowledge on ethical issues to bring adequate responses. We give a detailed example in section 3 to show how the approach could be implemented for a specific OR problem. We also discuss in section 4 how integrating practical ethics in tools (that are necessarily normative as they incorporate rules) could be beneficial in terms of promoting human agency.

2 Between abstraction and case-by-case approaches

In this section, we distinguish two opposite basic approaches to integrating ethics for an AI system in a concrete context: *principlism* and *on-the-ground*.

The first approach focuses on establishing multiple principles that an AI system should follow. One of the first instances of principlism was introduced by Beauchamp and Childress in 1979, who described four main ethical principles in the case of biomedical ethics: Beneficence, Non-maleficence, Autonomy, and Justice [4]. Since then, many institutions have produced different lists of ethical principles that should be followed or considered while designing an algorithm, whatever its nature. While these lists all differ from each other and may be large, they seem to converge and so can be summarized to a few basic principles [35, 40]. For instance, Floridi and Cowls reviewed different sets of principles produced by institutions to build an ethical framework, resulting in the inclusion of the four traditional biomedical ethics principles with the addition of Explicability [35]. Such sets of principles written by scholars and institutions aim to provide general guidelines for developers to design ethical (trustworthy, responsible...) AI systems.

In contrast, there exist on-the-ground or "case-by-case" approaches to integrating ethics into AI that do not necessarily rely on generic principles but primarily use practitioners' knowledge to arbitrate ethical issues. These approaches are typically used by developers who would like to question the ethical integrity of their work, having in mind business-relevant issues and goals [9].

2.1 Limitations of both approaches

2.1.1 Principles

Many sets of guidelines and principles for an ethical AI have been published by multiple entities, private or public. In 2019, Jobin et al. [40] identified 84 documents containing ethical principles or guidelines for AI from around the world. Some of these frameworks were focused on a particular field or published by governmental institutions, making them most relevant for the targeted contexts, but most were intended to be applied independently from field of application or location. In a specific situation, determining which framework would be the most appropriate in a specific situation may be a complicated task, as finding objective criteria for a relevant framework is not obvious. This accumulation of guidelines may also cause what Floridi called 'ethics shopping' [34]: picking the guidelines or principles that would 'retrofit some pre-existing behaviours', thus avoiding having to make changes that would be too inconvenient.

However, even if AI developers have chosen a single ethical framework to work with, there is still room for ethics shopping. Principles are meant to be generic so that they could be understood in any given context, meaning that their definitions may encompass many different notions that have little to do with each other. The notion of justice, a principle in 68 of the 84 documents reviewed in [40], is a great example of a single term encompassing a broad range of possible ethical issues. Guaranteeing justice may mean mitigating biases and ensuring fairness among groups of people in the AI outputs, giving opportunities to contest and appeal decisions, ensuring fair access to AI benefits, or even addressing societal issues. While all of these definitions differ largely from each other, they all are under the umbrella of 'Justice'. One could then 'shop' for their preferred definitions and ignore the rest of them, claiming they respect the Justice principle while their AI may be unfair on many points to different stakeholders.

Other obstacles like defining what should be considered as fair or how justice should be achieved make this concept even fuzzier. This means that any principle that is abstract enough could easily be implemented in the way that suits best, facilitating ethical 'box-ticking' [49].

Besides these risks of ethics shopping and box-ticking, translating ethical principles into concrete, on-the-ground situations is a great challenge in itself. Considering principles does not directly give a list of the practical problems to solve, but rather the way problems should be solved. When trying to interpret these principles and apply them to a real-life situation, one may ignore some of the issues, as principles do not give guidance or practical knowledge on the many particular issues that could be identified. For instance, the Justice principle applied in scheduling could be identified as ensuring a fair distribution of tasks only, ignoring that other distinct aspects may influence the fairness perception [18]. These guidelines might not give enough information to a developer, who is often not an expert of the specific field for which they provide algorithms, and may not be aware of the many problems their work could deal with. Without this knowledge, it is possible to overlook important aspects or to consider wrong assumptions while designing an algorithm, which means practical knowledge is necessary to provide ethical solutions. Principles have first to be interpreted into the given context, then it must be determined which tools should be used to deal with identified issues, and finally one implementation must be chosen and developed. Each of these questions must be answered by AI developers, and general ethical principles typically offer little help to answer them: as each situation is unique, a set of principles designed to be applied for every use case cannot provide these specific answers. It is then the developers' job to answer them, despite having probably little to no training on these topics [51] and no particular knowledge on the application field.

On top of these issues, algorithms may be used years after being designed by developers, so the consequences of decisions made by these algorithms might become more and more uncertain as time passes and the context in which they are implemented evolves. Taking the example of hospital wards, changing available equipment or having nurses perform new tasks affects the working conditions, meaning that a previously computed fair allocation of work could become unfair with these new conditions. Normative approaches such as principlism are unlikely to be sufficient to guarantee ethics in practice in the long term, which is why other approaches are needed.

2.1.2 On-the-ground

On-the-ground approaches are commonly used by experts to solve ethical issues and dilemmas day by day, by taking into consideration practical needs and expectations of stakeholders. These experts may not be specialised in ethics and may not rely on any academic literature or moral philosophy theories to make decisions [39]. At the same time, when they find a practical solution to a problem, they generally do not theorise about it nor make it accessible to the public. Companies are also typically less interested in ethics than other issues [65]. For these reasons, it could be challenging to find practical methodologies that are used in specific contexts.

While these practical approaches are usually straightforward and directly lead to action, they lack the systematic and comprehensive ethical bases that academics have

[9], potentially leading to some ethical concerns. For instance, developers' decisions may not be based on some criterion or ethical rules previously set, but rather solely on information they have on the situation they are confronted with and their personal judgement. Yet, their judgement might be unintentionally influenced by their internal biases and affects, which ethical frameworks may help compensate [56]. Case by case decision processes can also lead to a lack of unity or consistency, meaning several decisions may conflict with each other and can only be explained by ad hoc justifications, which also means that these decisions cannot be extended to another similar context. Such inconsistency can be perceived as unfair or inexplicable by stakeholders, because their situations do not undergo a clear and defined processing. Therefore, this kind of approach might not be sufficient as it does not follow systematic universal principles or values to drive decision-making, which can be criticised from an ethical standpoint. As decision-makers might be subject to a moral relativism [17], decisions that are supposed to be ethical may also be biased and driven by their company's needs, which might be ethically questionable in the first place. In this context, an 'ethical' decision might primarily benefit one stakeholder (the organisation) more than the other ones, which could be considered unfair.

Thus, on-the-ground approaches may be inefficient in many cases to deal with ethical issues as they are inconsistent and subject to various biases. On the other hand, ethical principles may be used as a basis for decision-making, but their generic aspect makes them insufficient in practice as they may easily be faulted by unforeseen circumstances where principles reach their limits and become unsuitable. Given the limitations of principlism and on-the-ground approaches, we propose another approach to decision-making based on academic field knowledge to help bridge the gap between principlism and on-the-ground approaches.

2.2 Using field-related knowledge

We argue for a more systematic and rigorous way to deal with ethical issues for a particular problem. One of the main issues related to principles is their abstract nature, which makes them hard to apply in practice. Indeed, they are thought to represent a Good that is not clearly defined. If we take a consequentialist perspective, which is common today when talking about the harm caused by AI [69], we need to be able to define the harm to be prevented. Principles alone are of limited help because they do not identify particular issues on the ground that can vary widely, and may imply very different practical solutions according to the case study. Therefore, we propose to first take into account issues coming from a specific field. By doing this, the goal is to be able to identify more accurately which ethical issues may be found within a particular field and details about their causes and consequences.

These issues may be identified directly from the field, as previously experienced harms may be detected or highlighted by stakeholders (e.g., unfairness in resource allocation) and show flaws in the decision-making process and/or the rules that are applied. When these issues are recurring, they may be studied by scholars, creating empirical and theoretical evidence in the target domain that can be drawn upon. Being able to pinpoint issues for a specific context allows for an efficient scientific literature

search on causes, consequences, and potential solutions for each issue. A necessary condition for this approach is that there exists a sufficient amount of scientific literature on the topic in question, otherwise this method might not help to identify appropriate solutions. We give an example of such a literature search in section 3.2 for health-related issues in shift scheduling problems. Then, if this process aims to directly deal with issues identified on the field in a specific context (e.g. a particular workplace), it will also be necessary to communicate with stakeholders to assess what points from the literature search might be most relevant to them and identify what changes are needed to current processes. This process may help with grasping the convolutions of a specific decision-making problem, as well as deliberating on which ethical principles should be prioritized when they cannot all be satisfied. From this point, the existing decision-making process must be questioned: are there problematic norms - as standards generated by automatic processing - that might directly or not cause harms? or, should new norms be added in order to correct flaws in the process? Involving all stakeholders in the design process makes sure that all needs and concerns are considered as early as the design phase, so that the challenges may be clearer. It must be noted that taking these inputs in consideration does not mean they will all be respected in the end, as the interests of many are likely to be conflicting.

Using knowledge and potential recommendations from specific research fields, it should be easier to understand what should be integrated into decision-making algorithms. Remaining work consists of translating this knowledge into mathematical tools. OR approaches for example, as model-based methods, are already used to explicitly integrate legal rules, preferences and so on [33]. Such an approach could therefore be enriched in multiple ways with additional standards, considering other stakeholders' inputs and recommendations from scientific literature.

By following these steps, it is possible to implement ethical technical solutions in concrete situations. Unlike a simple on-the-ground approach, ethical issues are addressed via systematic scientific study from which relevant decision rules can be derived. We will also demonstrate that this method may produce technical solutions that could be easily classified into the five principles that Floridi and Cowls [35] highlighted.

3 Including ethical considerations in algorithms: a case study for nurse scheduling

As we have argued, ethical analysis needs to get closer to a given identified object of study, in order to be able to instantiate principles, but also make use of dedicated theoretical knowledge, to avoid arbitrariness. Among all the ethical problems that are currently discussed around algorithms, we can find many of them related to work conditions and how AI in the larger sense is currently modifying them [25]. An emerging field or object of study, attracting lots of related discussions, has recently been named algorithmic management (AM), defined by Duggan et al. [29] as "a system of control where algorithms are given the responsibility for making and executing decisions affecting labour, thereby limiting human involvement and oversight of the labour process". This practice is increasingly used in industry, but also in many services such

as healthcare, and leads to a disruptive change in work organisation, with impact on many levels. Six main managerial functions have been identified by Parent-Rocheleau & Parker [53] that have been taken over by AM: monitoring, goal setting, performance management, scheduling, compensation, and job termination. Some companies have assigned the task of executing these functions to algorithms with an objective of increasing workers' productivity, but this mostly results in negative outcomes for workers in their activities [53].

Some of those negative effects could be linked with different works with older considerations around the effects of Automated Decision-Making Systems (ADMS), as an invisible technology [7] that provides "blind standards", invisible standards that are by definition not necessarily discussed or even stated. Given current awareness of the creation of standards in ADMS, an idea could be to study how those quite invisible standards could in fact be made visible, discussed, and thought of, in order to enrich the ADMS with some desired ethical norms. In that context, this paper will now focus on one of the particular managerial functions just cited:work scheduling. In that smaller field, it is possible to consider numerous specific questions and a dedicated literature in computer science.

3.1 Problem overview

In this section, we will introduce the specific case we focus on in order to illustrate how it could be possible to integrate norms into a given ADMS. We choose for this part a problem of shift scheduling, which means allocating workers to different shifts, so that demand is covered for each period of time. This system is widely used for different types of workers, including factory workers, nurses, and telemarketers. In order to take into account more specific work related issues, we focus on nurse scheduling, which is a widely studied problem both in Operations Research, where it is known as the Nurse Rostering Problem (NRP) [23], as well as in human sciences [6]. The NRP is a particular case of shift scheduling, and consists of assigning nurses to shifts to be worked for a given period, assuming demand for each shift is known as well as the set of possible shifts (starting times and durations). Therefore, this problem is dependent on higher level decision problems like dimensioning, shift design or demand estimation and also impacts lower level problems such as rescheduling [38]. Particular rules and criteria might be added on top of this basis to fit field reality. The context of our study will consider the integration of a scheduling ADMS into a hospital department, especially focusing on ethical issues regarding the personnel.

Framing the topic this way carries many advantages for three main reasons. First, health and well-being issues for shift workers and/or nurses is a subject that has been studied extensively by scholars, who have analysed multiple situations involving these types of workers and have made general ergonomic recommendations, including schedule recommendations. Using this research is fundamental if we would like to limit negative effects on workers' well-being. Second, we place ourselves in a specific context where main stakeholders are easily defined (human resources, nurse managers, nurses, patients), and work activities are mostly defined (they could still vary from one department/hospital to another). While not targeting a particular real-life context, we may have a clear idea of the environment of stakeholders, the interactions between

them, and how the introduction of an ADMS may cause disruptions, as it is also a well-studied topic. Third, studying a problem that is already well researched in the technical literature gives us a solid basis on which we can integrate new rules and recommendations.

While we consider here only the impact of schedules on nurses, patients' well-being may also be directly impacted by these decisions in different ways. For example, if a nurse manager decides to prioritise patients over nurses, then nurses may be required to work longer or during less suitable hours to ensure patients are treated correctly. Nurses themselves may also be willing to work more hours for the sake of patients, thus making ethical decisions at their own level. To clarify the presentation of our method, we chose to limit the study to solely scheduling decisions and considerations regarding nurses' well-being, ignoring other concurrent ethical issues that may present themselves.

Given this context, we would like to integrate the findings from many studies on the subject into technical tools. A literature review can help identify important issues and recommendations, and which ethical principles they relate to. This integration would constitute a first normative step, as we would take into account field recommendations to design an ADMS, performing a stronger ethical analysis to form acceptable solutions or sort them according to relevant criteria. In the next section, we will cover reviews and studies that are of interest in our case, especially related to occupational health. We will then highlight specific recommendations that have been produced, and afterwards, we will propose ways to translate these recommendations into technical and workable elements to expand or reinforce existing models, making them more suitable for an ethical processing.

3.2 Harms and recommendations found in literature

Shift work is defined by Knutsson [44] as "a work schedule in which a worker replaces another on the same job within a 24 h period". Many workplaces are using shift work, as it allows for a continuous service or production, while balancing work hours between employees. However, the constant change of pace and regular night work that are implied by this system may cause various troubles for the workers, directly affecting their health and well-being.

One of the main issues regarding shift work is the sleep problems that workers experience [2]. Due to the regular changes of routine that are implied by this system, their body may have issues adapting to different hours of activity and sleep, leading to circadian rhythm disruption. Misalignment of this rhythm may cause 'jet lag' syndrome, characterised by fatigue, sleepiness, insomnia and irritability. Furthermore, workers having to sleep during the daytime (typically after a night shift) usually sleep less due to their metabolism reaching maximum alertness [1] and having to sleep in an inappropriate environment with disturbing lighting and noises [20]. Before morning shifts starting too early, sleep time may also be reduced due to workers' social life having them sleep later than recommended. This reduced sleep can lead to tiredness during work, and may cause fatigue in the long-term [20]. Using a forward rotation system (e.g. morning shift followed by an evening shift followed by a night shift) instead

of a backward one (e.g. night shift followed by an evening shift followed by a morning shift), as well as avoiding particular sequences may help reduce circadian rhythm disruption [42, 66].

Circadian rhythm disruption may cause other short term health disorders like gastrointestinal malfunctions due to 'jet lag', but also long term ones. Knutsson [43] identified that ulcers, cardiovascular diseases or pregnancy outcomes (miscarriage, low birth weight, preterm birth) are linked with shift work. Moreover, existing disorders might be aggravated as effectiveness of medications can be modified due to biological clock desynchronization and sleep deprivation. Long working hours are also a factor of most of these diseases as well as some psychological conditions [24, 59]. Aside from errors that could happen during work time due to tiredness (which can impact patients' health in our case), likelihood of drowsiness while commuting home is increased for shift workers, especially after a night shift [27].

Shift work also has an impact on social life as timetables are often mismatched with family and social activities, which are mostly planned according to a day-oriented rhythm. Shift workers often have troubles maintaining a good work-life balance, which may lead to social marginalisation and worsen marital relationships, parental roles, and children's education [19]. This is also due to the fact that shift workers may have only few control over their schedule, which makes it unpredictable [21].

Across different fields, shift workers may face similar consequences, regardless of their specific jobs. Nevertheless, depending on their specific job activity and environment, some generic rules that could be applied regarding well-being improvement might not be relevant for everyone. For instance, implementing rules for limiting night work would be one of the best ways to limit the many consequences [42], but some services cannot reduce their staffing at night for safety reasons that were decided beforehand. An understaffed hospital ward may have a direct impact on both patients' health [37] and on-duty nurses, whose workload would greatly increase. Even between same-activity organisations, elements like culture, tasks, and personalities may affect a norm's perception and efficiency. Even if a holistic study of a specific workplace might be unreasonable, taking into account these elements may help determine a norm's relevancy in a particular workplace. This can be achieved through discussions and feedback from stakeholders, which could also help with identifying new specific rules relevant for the particular workplace.

All of these potential issues may directly affect shift workers' health and well-being. An AI whose role would be to design shift schedules should integrate these concerns into its decision process, and not just legal working rules as per standard approaches. From an ethical standpoint, taking into account principles highlighted by Floridi and Cowls [35], this integration falls under the umbrella of Beneficence, 'promoting the well-being of people and the planet' and Non-maleficence, requiring AI to 'do no harm' ([35] p. 6). In the current context, addressing the latter principle could mean ensuring that the AI does not create schedules that would put people into a harmful condition, while the former principle means the AI should provide schedules that actively benefit workers' lives. In the same context, one could identify issues that relate to other

principles than Beneficence and Non-maleficence, such as ensuring schedules are fair for every nurse would relate more to the Justice principle.

Of course, in order to ensure these principles are respected within a workplace, other aspects than staff scheduling must be taken into consideration such as other human resources management functions or work activities. As positive a schedule can be for workers' well-being, ignoring these other important facets may diminish its benefits, or just make them insufficient to guarantee an acceptable level of welfare.

Using literature on psycho-social risks that can be faced due to scheduling circumstances, it is possible to build mathematical constraints that could be used to represent them, in order to avoid identified negative impacts on workers. For instance, we take into account the nine recommendations for the organization of shift schedules presented in the review by Costa [20]: a) limit night work as much as possible; b) avoid a large number of consecutive night shifts; c) prefer quickly rotating (every 1-3 days) shift systems to slowly rotating (i.e. weekly or longer) ones and to permanent night work; d) prefer clockwise rotation (morning/afternoon/night) to the counter-clockwise (afternoon/morning/night) rotation; e) set the length of shifts according to psycho-physical demands; f) avoid morning shifts that start too early; g) set an adequate number of rest days between shifts, particularly after night shifts; h) keep the shift system as regular as possible; i) allow flexible working time arrangements according to worker's needs and preferences. As the targeted problem concerns shift allocation, some of these recommendations are irrelevant at this level of decision and must be decided in a shift design phase (length and start of shifts) or at strategic level (limit night work). Nevertheless, these decisions will then have an impact on the shift allocation problem and may or may not facilitate the design of an 'ethical' schedule.

3.3 Integrating recommendations in a model

There are multiple ways to deal with these potential negative effects in a technical approach. In the case of scheduling, applying OR methods is very common [14] as it is an effective way to represent the multiple rules that must be taken into account regarding hours of work. These methods often consist of constraints and single/multiple objective function(s). While hard constraints are a mathematical representation of a set of rules that have to be respected (mainly due to work regulation), objective functions and soft constraints are used to attribute values to each schedule, allowing comparison with each other, in order to determine the supposed best solution(s). Mathematically, hard constraints consist of inequations used to delimit the space where the acceptable solutions are located, whereas soft constraints do not have to be respected but a solution not respecting them will be penalized increasingly as the solution deviates more from soft constraints. Implementing soft constraints instead of hard ones may help find more feasible solutions in the end.

A direct way to integrate new ethical norms into a mathematical model would be to translate them into mathematical constraints, potentially using new variables, and simply append them to the model. Choosing this approach can be compared to deontological ethics, as it creates a set of ethical rules that should always be respected. If instead we want to integrate a differentiation criterion, norms could be translated

into an objective function, or multiple objective functions. Using ethical criteria that should be optimised may relate more closely to utilitarian approaches in this case. Some standards may not be directly mathematically translated and integrated into a model, but may require ADMS design changes, as we will explain later.

Consecutive nights and rest days

Some recommendations include avoiding a large number of consecutive nights and allowing an adequate number of rest days between shifts. These recommendations correspond to individual workers' schedules and may be directly translated into 'hard' constraints, prohibiting schedules that do not respect thresholds for each worker. There may not be many ways to express mathematically these kinds of constraints, but a main difficulty concerns the way of implementing thresholds. For instance, number of consecutive nights shifts could be limited to a fixed number (e.g. three, according to Knauth and Hornberger's recommendations [42]). However, this limit may be considered too high or too low depending on context (worker's condition, nature of work activities, workload, previous worked shifts, level of staffing...), so that a fixed threshold would not be suitable. There may typically be a need to formulate new rules to determine how many consecutive worked night shifts would be acceptable in different contexts (e.g. two when workload is considered high, three when considered low), resulting in new constraints to implement. In addition to hard constraints, soft constraints may be added to represent ideal thresholds that may not be respected, but any solution including a deviation from it would be penalized to the extent of the deviation [36].

Sequences

Particular sequences of shifts may also be preferable to others. Shift systems including quick rotations of same-type shifts (e.g. three shifts of type A in a row, followed by three shifts of type B in a row...) seem to have least impact on workers' health than slowly rotating systems, with same-type shift sequences of a week or more [66]. Similarly, clockwise rotations (Morning $\rightarrow$ Evening $\rightarrow$ Night) help limit problems of adaptation of circadian rhythms as opposed to counter clockwise rotations (Night $\rightarrow$ Evening $\rightarrow$ Morning) [66]. Some identified particular sequences are also not recommended because of their impact on workers' well-being [42]. All of these sequences can be taken into account by restricting them or allowing only a subset of sequences in the model constraints [12]. As previously said, some of these constraints might be suitable or not depending on context, person, and so on, and can be adapted accordingly by targeting relevant groups of workers only.

Shift system regularity

Another recommendation concerns the regularity of the shift system, which can be defined as the variability in schedules from one period to another [50]. From a worker perspective, keeping the shift system as regular as possible may increase its predictability and avoid work-life conflicts [10]. Mathematically, the distance between a basic schedule pattern and a particular schedule may be measured in different ways, one of the simplest being computing how many shifts differ from the pattern. Then,

this value can be constrained using a set threshold (e.g. each worker must not have more than six shifts differing from the pattern) and added to the constraint list. Another solution to keep schedules regular would be to integrate workers' preferences on working/non-working days that should stay fixed as much as possible, in order to facilitate their work-life balance management. Of course, conflicts may arise between workers' demands on the same days, and would have to be arbitrated technically (e.g. using fairness measures in the mathematical model [68]) or otherwise (e.g. interpersonal resolution [63]).

Flexible working time arrangements

More generally, workers' work-life balance may profit from flexible working time arrangements. Here, flexibility is dissociated from variability: the former defines autonomy and individual discretion regarding one's schedule, the latter defines employer control over it [21]. Implementing into an ADMS ways for workers to input their own preferences for the next period, either individually set or coming from collective arrangement and discussions, may increase flexibility. The nature and number of these preferences have to be decided, as well as how they should be allocated in case of conflicts.

Ideally, all of the above recommendations could be implemented in a model. However, not all recommendations might be relevant for a given team of nurses and some of the recommendations might not be compatible with each other, which can cause issues with problem resolution. By adding new hard constraints into a model, the set of feasible solutions shrinks and may become empty. Even by considering this potential problem and only adding a subset of these rules, there might still be some incompatibilities for some periods where a choice has to be made about which rule(s) will not be respected. For instance, if a hospital ward is understaffed on a certain period, constraints for nurses' well-being might not be respected if demand needs to be met each day. In order to get a valid schedule, one or several of these constraints should be relaxed. Another example would be deciding which nurse should have their preferences satisfied where several nurses have the same request. In these cases where an ethical dilemma emerges, a careful choice has to be made on what rule should be overridden.

By implementing all or some of these rules into an existing mathematical model, new ethical norms and/or criteria are created, making schedules more acceptable in regards to them. In the general case, changing a model by adding, removing or changing parts of it has an influence on its intrinsic values and norms. Whereas improving nurses' health and well-being could be categorised as helping to address 'Beneficence' and 'Non-maleficence' principles, other types of issues in decision-making processes may be resolved by adding new specific norms into technical tools. Nevertheless, these rule-based tools may be too complex in practice, meaning nurses may not use them at all [28]. In addition, some ethical issues in ADMS cannot be solved using only normative tools, such as mathematical constraints, as we will discuss in the next section.

4 Addressing limitations of automated decision-making tools

4.1 Harms coming from autonomous decision-making system

Philosopher Bernard Stiegler uses the term of 'digital pharmakon' [61], echoing Socrates, to capture the two sides of any algorithmic processing : Socrates showed in Phaedrus [55] that writing, as any technique, is a pharmakon, meaning it is fundamentally ambivalent, both a remedy and a poison at the same time. B.Stiegler uses the term 'digital pharmakon' to highlight the immense benefits, but also the intrinsically linked threats, of algorithms and related services. In decision-making for instance, algorithms both help people to compute huge amount of data with advanced mathematical tools and at the same time poison our abilities to have an autonomous way of thinking and to analyse the situation to make a decision. Because the decision relies partly on these algorithms, one becomes more vulnerable to be influenced by a partially-oriented optimisation, just as large language models are now known to create plausible - but false - facts [8]. Following Stiegler's assumption, we can infer that even with a thorough research and algorithmic implementation of norms aiming to prevent unethical issues from happening, there will still be issues that cannot be resolved this way. We identify here three main issues inherent to ADMS.

First, any kind of automation people are trying to develop in order to improve speed and avoid deviation of a given process ideally leads to exactly what it is designed for: automatically and quickly reproducing the captured process, on a very large scale. On that basis, errors, deviations and instabilities of a hand-made, that is to say non-automated, process should be avoided. But unfortunately, the process that has been captured to conceive an algorithm is never a complete representation. One may find some particular circumstances where experts would not follow that given process and make an exception. Human ability to reframe a problem as they try to solve it "on the ground" leads to an infinite number of adaptations that the algorithm could not achieve [3]. Indeed, it is not possible for an algorithm, that is not based on meaning but on representation, to suddenly integrate a new type of data, question previous decisions, redefine the objective, or include ethical considerations that have not been specified and incorporated, even if they apply in the current situation. This would then result in a mishandling of cases where particular adaptation is required, potentially causing harm, depending on the situation [15].

Besides the people handled by an ADMS experiencing ethical issues (e.g. nurses), ethical issues can also impact the people working directly with the ADMS, which we call here decision-makers (DM). Depending on the level of automation of the ADMS they use, a DM will more or less lose some of their autonomy, delegating tasks to the system. Yet, autonomy is one of the five principles highlighted by Floridi and Cowls [35], in the sense of promoting human agency and being able to be in control of the decision-making process. For instance, a DM should be able to arbitrate whenever decision-making tasks can be automated and to revert a decision made by automation [26].

But, in order to understand whether an automated decision should be reverted or not, one needs to have a good appreciation of the situation. If a DM is unable

to detect flaws in the decision, then it does not matter that they are able to revert it. In the same way, if there is a malfunction in the system, the DM should be able to perform manually the malfunctioning tasks. Unfortunately, in multiple instances, researchers have shown that automation may cause loss of situation awareness and skill degradation, among other things [22, 31, 32, 54].

4.2 Balancing human autonomy and automation in human-AI collaboration

As automation itself causes issues impacting multiple stakeholders, methods have been developed to identify in a system the nature of automated tasks and methods to keep the human at the center of the decision process. In order to identify how much a system is automated, descriptive tools for measuring levels of automation have been developed. Vagia et al. [64] reviewed taxonomies describing levels of automation of systems designed for different applications. Most of them range from manual to fully automated tasks and include notions of decision support, human approval, and veto to describe different levels. Sheridan and Verplank [58] were the first to introduce multiple levels of automation with a scale of ten different levels of cooperation between a human operator and a computer. Bruni et al. [13] introduced a Human-Automation Collaboration Taxonomy which identifies distinct stages and three roles in decision-making processes: moderator, generator and decider. Moderator is the agent keeping the process moving forward, Generator is the agent generating solutions and Decider is the agent making the final decision. For any decision-making process, these three roles are given a score on a scale of 1 to 5, 1 meaning the role is assumed by human only, 5 meaning the role is assumed by automation only, and 2 to 4 if they both share the role, equally or not.

These taxonomies highlight trade-offs between system automation and human autonomy. The lower the level of automation, the more likely the human detection of harms or decision flaws before the decision is made. At the same time, performing manual tasks only seems inefficient, as fast computation may substantially reduce tasks' length, which might profit organisations as well as decision-makers. Thus, in order to both have a satisfying decision process length and allow for enough human autonomy, a good human-automation trade-off has to be found. As highlighted in [13], different kinds of tasks can be performed by either the human or machine agent, and some of these might be more appropriate for one or another. For instance, Decider tasks might require more human involvement than the Generator ones, where automation is much more efficient than humans and there are no critical decision tasks. Another possible framework to preserve autonomy is called 'meta-autonomy' by Floridi and Cowls [35], where human agents decide by themselves which decisions they should make and which ones they cede to automation.

In order to assess if a DM in a given situation has enough agency to make appropriate changes, three questions could be asked. First, do they have the competency to understand solutions and their potential consequences to amend them accordingly? Second, does the decision-making system integrate ways for the human agent to make these amendments? Third, does the system give sufficiently good explanations on

solution characteristics and help them understand the problem structure?

The first point questions the training and skills of a DM regarding a decision-making task, as well as their situation awareness and practical knowledge. A DM should have a minimum amount of information regarding the problem beforehand, whether by training or experience, in order to at least detect special cases or unwanted patterns, for instance. Situation awareness can be defined as 'an internalized mental model of the current state of the operator's environment' [30] and is crucial when it comes to decision-making. Without correct information on their environment, a DM might make unethical decisions, such as not considering workers' characteristics and needs when designing schedules. For instance, if a nurse manager is unaware of the individual preferences regarding days off, they would not be able to integrate this information in the decision making process, which would result in unsatisfying schedules. Such issue is not resulting from the scheduling process itself but happened because of the DM's lack of information regarding the situation. Nonetheless, data collection of preferences could be integrated within the whole scheduling process to make sure the nurses' preferences are considered by the DM, which is something that is already done more or less formally in most hospitals.

The second point directly concerns autonomy and the DM's control over the process. If a system does not integrate human inputs into the decision loop, the DM might have to circumvent the system to handle special cases, which might be time consuming and cause organizational dysfunctions. Integrating human in the loop does not necessarily solve this problem if control is too limited, such as immutable fields or limited options. In the case of scheduling, it could be possible to obtain an invalid schedule due to outdated constraints or blind spots of the model implementation. If the scheduling tool does not allow for user interaction, the DM might have to find workarounds to obtain a valid schedule, which would most likely imply some manual scheduling. By allowing the DM to change the rules, such unforeseen issues are more likely to be easily circumvented and does not prevent using the advantages that automation provides.

Finally, the third point refers to explicability, especially including intelligibility and transparency aspects. In order to effectively modify or create new solutions, one must be aware of the possibilities offered by the system. For complex decision-making tasks like scheduling that involves combinatorial problems, modifying tiny parts of a valid solution can make it unfeasible or greatly worsen its quality. If the ADMS does not give explanations or at least warn about consequences of wanted changes, it can result in unforeseen situations. By helping the human agent to understand the problem in real time and eventually giving them recommendations, it might improve their experience and facilitate their decision-making. As the decision-making is a hybrid task in this case, prompting intelligible details on the automated parts to give more insights on the problem solving may help with the reasoning.

In the next section, we give a preview of a potential decision-making process for nurse scheduling that helps guarantee these three points.

4.3 Towards efficient human-AI collaboration in nurse scheduling applications

Korhonen [45] defines Multi-Criteria Decision Making (MCDM) as decision and planning problems involving multiple (generally conflicting) criteria. In the field of MCDM, interactive methods are used to identify which solution is preferred by a DM through interactions with the optimisation system. There are different approaches to interactive methods [45], but they all share the same core steps, which are: 1) Initial solution(s), 2) Evaluation, 3) Solution(s) generation, 4) Termination. In step 1, one or many solutions are displayed to the decision-maker, with corresponding criteria values. In step 2, the DM provides their preferences to the system regarding these solutions and criteria. In step 3, new solution(s) are generated according to the given preferences. These two last steps are repeated until the decision-maker decides to stop the process (step 4). This approach allows decision-makers to explore (a subset of) the solutions space to understand which trade offs between the multiple criteria exist. The system may also help the user to have a general view over the many possibilities and provide different information about the solution space. As multi-criteria problems can be complex to fully grasp, especially when there are more than two criteria, this method focuses on helping the decision-maker to find a good balance between various criteria, without specific consideration of the underlying solution. The initial problem and criteria formulation are also supposed correctly specified in the mathematical model.

We propose here an adaptation of the interactive methods framework to design decision-making tools that take ethical aspects into account. We will focus here on its application on an NRP, but other types of problems solved with Operations Research models may also be relevant. Our main goal is to create efficient ways for a DM to manipulate a model as they desire, by changing it and exploring its potential solutions and associated ethical impacts. This way, we hope to give DMs support for a better understanding of the problem and enough agency to allow them to be flexible in the face of unforeseen contexts. First, we want to start the process from a 'core' model, containing strictly functional and eventually legal constraints. This model is assumed to be an unbiased one as it does not take in consideration any preference from the decision-maker or workers, and contains only rules that must be respected in theory. Considering the core model, a feasible solution is targeted (in this case, displayed as a timetable to the DM).

If such a solution does not exist, meaning the set of feasible solutions is empty, some of the basic constraints must be relaxed in order to obtain a solution. This situation may happen in the case of understaffing, where it is not possible to cover demand for the whole period as well as respecting legal working hours rules. In this case, the DM may change the model by relaxing one or multiple constraints in order to obtain at least one feasible solution, just as constraints are sometimes relaxed in real-life cases, always relying on humans. System recommendations can be displayed to suggest some less harmful relaxations to obtain a non-empty feasible set.

When a solution is displayed, the DM should be as free as possible to modify it in order to explore the set of feasible solutions by interchanging elements of the schedule, or making the optimiser generate new solutions according to chosen criteria, to get more suitable solutions. When potentially problematic patterns are automatically

detected in a solution, a flag is raised to indicate potential flaws and the related ethical considerations that could be avoided by adding new rules. These patterns could be derived from normative rules found in the literature as highlighted in section 3.2. The DM can then choose to ignore these signals if they are irrelevant in their opinion, or to add the constraints related to the recommendation to avoid some situation. At the same time, it should be possible for them to integrate new constraints for specific cases, such as when one particular nurse should not undergo a certain pattern; or more generally, when no nurse should undergo this pattern. After such modification, a new feasible solution according to the changes is computed and displayed. Then, the DM can make changes again to this solution and the process loops until they are satisfied with a solution and no change has to be made.

This human-in-the-loop process allows the DM to be in full control, by giving them tools to create their own solution. The computation power of the system is mostly used to generate new solutions according to decisions made by a human agent, who ultimately decides what is the best solution according to their constraints. With help from system recommendations, the process helps ensure a good level of human autonomy in the decision-making, as compared to system automation. It also guarantees a certain level of transparency and explicability, with counterfactual reasoning and problem explanations to empower users in their decision-making abilities. Its flexibility also ensures there might be possibilities for the decision-maker to input particular constraints that were not formalised beforehand. Nevertheless, user input is still limited by the set of possibilities they are given. If there are very specific situations that the DM wants to represent that were not thought of beforehand by developers, it might be impossible for them to convey the desired adjustments to the system. This potential problem means the designing phase of such tool should be as thorough as possible and include as many stakeholders as possible to help avoid such a situation.

5 Conclusion

Principle-based approaches in AI ethics suffer from multiple flaws that make principles hard to apply in specific contexts. At the same time, relying solely on decision-makers' field experience or AI developers' efforts to arbitrate ethical dilemmas might be inefficient and inconsistent, as they may not rely on universal values to drive decisions. Helping to bridge principles and practice, we argue for a way to integrate ethics into decision-making algorithms by considering field-related knowledge. According to specific needs, a literature search can be performed to retrieve key elements such as main causes and consequences of potential ethical issues, and derive implementable rules. As an example, we took into consideration well-being issues in a hospital's continuous service, functioning with staff shifts. Literature showed that shift work might have a negative impact on multiple well-being aspects, some of which can be mitigated by integrating specific rules in a schedule design process. These rules could be translated into mathematical tools such as objective functions or constraints, and added to an existing mathematical rostering model. While not directly relying on abstract ethical principles, dealing with such issues could easily be considered as applying Beneficence

or Non-maleficence for a specific class of workers, but other principles could be similarly addressed using the same method. For example, the Justice principle could also be addressed by considering findings in organisational justice. Basing mathematical tools on a scientific literature review gives a solid justification that could be more accepted and beneficial than arbitrary decisions.

However, simply adding new norms this way presents some limits. Specific or newly arising cases may appear, for which previously set norms are not relevant or incompatible. Also, some ethical rules can contradict each other and an arbitration, necessarily a human one (as they are the only possible moral agent [41]), has to be made in order to obtain a solution. For these reasons, decision-maker autonomy is a critical point that must be considered in order to ensure an appropriate system. This feature translates technically by giving the DM tools to interact with the system as well as relevant explanations on problem structure and feasible solutions.

Thus, while not using a principle-based approach, we have proposed a method for operationalizing all five principles of Floridi and Cowls' framework of AI ethics [35] for decision-making algorithms. First, conducting a literature search to implement new rules within an ADMS can be helpful to comply with Beneficence, Non-maleficence or Justice depending on the nature of these rules and ethical issues they are supposed to deal with. For instance, norms in nurse schedules can be used to either improve nurses' well-being or to avoid unhealthy work patterns, thus contributing to Beneficence or Non-maleficence. Meanwhile, the Justice principle could be respected at the scheduling level by ensuring a fair allocation of shifts among nurses (distributive justice), or by designing tools that allow for nurses' input during the process (procedural justice). Going further, using a fully autonomous tool to perform these tasks decreases human autonomy. Allowing more interactions between humans and the system while also enhancing system Explicability ensures that a DM acts in good Autonomy. As the whole process may be more transparent and acceptable for workers as well, this could also contribute to procedural justice.

While this method might not be suitable for all problems, the specific case we chose seems promising. Experiments with decision-makers and nurse teams are needed to test how efficiently a system allows the integration of ethical considerations into the decision process and what could be the impact of integrating other common criteria, such as productivity or cost [23]. We may assume that they are conflicting objectives in many cases, but ethical aspects still have to be taken into account for many reasons. Another direction to consider will be to integrate interactions, such as preferences input, directly coming from stakeholders in order to go further in taking all points of view into account. Designing adequate interactions with dedicated scheduling tools could help us highlight helping and limiting mechanisms in regards to ethical considerations by analyzing if users change their decision while using such tools. As users may not easily spot ethical issues in staff scheduling problems due to their complexity, we could also incorporate mechanisms to highlight potential problematic situations in a given schedule. These steps could be significant for improving the design of scheduling tools, especially if experimental evidence support this kind of approach.

This method is not intended to be a general one, as a given AI tool needs to integrate dedicated literature, which obviously does not exist for any issues yet to come.

Additionally, any normative tool still has to be refined, when facing additional problems along with previously identified challenges. Our concept continues OR debates that started in the 70's on ethical decision-making [11], trying to integrate current knowledge. There still exist some issues we do not have in mind and do not study yet, so we obviously do not know how to balance them. This approach will have to be extended further, but will hopefully be a significant step in ethics integration in decision-making tools.

References

[1] Åkerstedt T (2003) Shift work and disturbed sleep/wakefulness. Occupational medicine 53(2):89–94

[2] Åkerstedt T, Wright KP (2009) Sleep loss and fatigue in shift work and shift work disorder. Sleep medicine clinics 4(2):257–271

[3] Barthélemy J, Bisdorff R, Coppin G (2002) Human centered processes and decision support systems. European Journal of Operational Research 136(2):233–252. https://doi.org/https://doi.org/10.1016/S0377-2217(01)00112-6

[4] Beauchamp TL, Childress JF (2001) Principles of biomedical ethics. Oxford University Press, USA

[5] Bellenguez O, Brauner N, Tsoukiàs A (2023) Is there an ethical operational research practice? and what this implies for our research? EURO Journal on Decision Processes 11:100029. https://doi.org/doi.org/10.1016/j.ejdp.2023.100029

[6] Belorgey N (2010) L'hôpital sous pression. La Découverte

[7] Berry M (1983) Une technologie invisible - L'impact des instruments de gestion sur l'évolution des systèmes humains. Cahier du laboratoire, classification JEL : L20

[8] Bian N, Liu P, Han X, et al (2023) A drop of ink may make a million think: The spread of false information in large language models. arXiv preprint arXiv:230504812

[9] Blackman R (2020) A practical guide to building ethical ai. Harvard Business Review 15

[10] Bohle P, Quinlan M, Kennedy D, et al (2004) Working hours, work-life conflict and health in precarious and" permanent" employment. Revista de saúde pública 38:19–25

[11] Brans JP, Gallo G (2007) Ethics in or/ms: past, present and future. Annals of Operations Research 153:165–178

[12] Brucker P, Burke E, Curtois T, et al (2010) Adaptive construction of nurse schedules: A shift sequence based approach. Journal of Heuristics 16(4):559–573

[13] Bruni S, Marquez JJ, Brzezinski A, et al (2007) Introducing a human-automation collaboration taxonomy (hact) in command and control decision-support systems. In: 12th International Command and Control Research and Technology Symposium, Command & Control Research Program Newport, MA, pp 1–13

[14] Burke EK, De Causmaecker P, Berghe GV, et al (2004) The state of the art of nurse rostering. Journal of scheduling 7:441–499

[15] Calvo RA, Peters D, Vold K, et al (2020) Supporting human autonomy in ai systems: A framework for ethical enquiry. Ethics of digital well-being: A multidisciplinary approach pp 31–54

[16] Cath C (2018) Governing artificial intelligence: ethical, legal and technical opportunities and challenges. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 376(2133):20180080

[17] Churchman CW (1970) Operations research as a profession. Management science 17(2):B–37

[18] Colquitt JA, Greenberg J, Greenberg J (2003) Organizational justice: A fair assessment of the state of the literature. Organizational behavior: The state of the science pp 159–200

[19] Costa G (2003) Shift work and occupational medicine: an overview. Occupational medicine 53(2):83–88

[20] Costa G (2010) Shift work and health: current problems and preventive actions. Safety and health at Work 1(2):112–123

[21] Costa G, Sartori S, Åkerstedt T (2006) Influence of flexibility and variability of working hours on health and well-being. Chronobiology international 23(6):1125–1137

[22] Cummings M (2004) Automation bias in intelligent time critical decision support systems. In: AIAA 1st intelligent systems technical conference, p 6313

[23] De Causmaecker P, Vanden Berghe G (2011) A categorisation of nurse rostering problems. Journal of Scheduling 14:3–16

[24] Dembe AE (2009) Ethical issues relating to the health effects of long working hours. Journal of Business Ethics 84:195–208

[25] Deranty JP, Corbin T (2022) Artificial intelligence and work: a critical review of recent research from the social sciences. AI & SOCIETY pp 1–17

[26] Dignum V (2017) Responsible autonomy. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17, pp 4698–4704, https://doi.org/10.24963/ijcai.2017/655

[27] Dorrian J, Tolley C, Lamond N, et al (2008) Sleep and errors in a group of australian hospital nurses at work and during the commute. Applied ergonomics 39(5):605–613

[28] Drake R (2019) Dilemmas of e-rostering old and new: towards intelligent systems. Nursing Times 115(6):19–23

[29] Duggan J, Sherman U, Carbery R, et al (2020) Algorithmic management and app-work in the gig economy: A research agenda for employment relations and hrm. Human Resource Management Journal 30(1):114–132

[30] Endsley MR, Jones W (2013) Situation awareness. The Oxford handbook of cognitive engineering 1:88–108

[31] Endsley MR, Kaber DB (1999) Level of automation effects on performance, situation awareness and workload in a dynamic control task. Ergonomics 42(3):462–492

[32] Endsley MR, Kiris EO (1995) The out-of-the-loop performance problem and level of control in automation. Human Factors 37(2):381–394. https://doi.org/10.1518/001872095779064555

[33] Ernst A, Jiang H, Krishnamoorthy M, et al (2004) Staff scheduling and rostering: A review of applications, methods and models. European Journal of Operational Research 153(1):3–27. https://doi.org/https://doi.org/10.1016/S0377-2217(03)00095-X

[34] Floridi L (2019) Translating principles into practices of digital ethics: Five risks of being unethical. Philosophy & Technology 32(2):185–193. URL https://doi.org/10.1007/s13347-019-00354-x

[35] Floridi L, Cowls J (2019) A Unified Framework of Five Principles for AI in Society. Harvard Data Science Review 1(1)

[36] Glass CA, Knight RA (2010) The nurse rostering problem: A critical appraisal of the problem structure. European Journal of Operational Research 202(2):379–389

[37] Glette MK, Aase K, Wiig S (2017) The relationship between understaffing of nurses and patient safety in hospitals—a literature review with thematic analysis. Open Journal of Nursing 7(12):1387–1429

[38] Hulshof PJ, Kortbeek N, Boucherie RJ, et al (2012) Taxonomic classification of planning decisions in health care: a structured review of the state of the art in or/ms. Health systems 1:129–175

[39] Ibáñez JC, Olmeda MV (2022) Operationalising ai ethics: how are companies bridging the gap between practice and principles? an exploratory study. AI & SOCIETY 37(4):1663–1687

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

[41] Jonas H (1984) The imperative of responsibility: In search of an ethics for the technological age. University of Chicago press

[42] Knauth P, Hornberger S (2003) Preventive and compensatory measures for shift workers. Occupational medicine 53(2):109–116

[43] Knutsson A (2003) Health disorders of shift workers. Occupational medicine 53(2):103–108

[44] Knutsson A (2004) Methodological aspects of shift-work research. Chronobiology international 21(6):1037–1047

[45] Korhonen P (2005) Interactive methods. Multiple criteria decision analysis: state of the art surveys pp 641–661

[46] Madiega T (2021) Artificial intelligence act. European Parliament: European Parliamentary Research Service

[47] Mittelstadt B (2019) Principles alone cannot guarantee ethical ai. Nature machine intelligence 1(11):501–507

[48] Mittelstadt BD, Allo P, Taddeo M, et al (2016) The ethics of algorithms: Mapping the debate. Big Data & Society 3(2). https://doi.org/10.1177/2053951716679679

[49] Munn L (2019) The uselessness of AI ethics. AI and Ethics https://doi.org/10.1007/s43681-022-00209-w, URL https://doi.org/10.1007/s43681-022-00209-w

[50] Nabe-Nielsen K, Garde AH, Aust B, et al (2012) Increasing work-time influence: consequences for flexibility, variability, regularity and predictability. Ergonomics 55(4):440–449

[51] Oliver JC, McNeil T (2019) Undergraduate data science degrees emphasize computer science and statistics but fall short in ethics training and domain-specific context. PeerJ Computer Science 7. https://doi.org/10.7717/peerj-cs.441, publisher: PeerJ Inc.

[52] O'Neil C (2017) Weapons of math destruction: How big data increases inequality and threatens democracy. Crown

[53] Parent-Rocheleau X, Parker S (2021) Algorithms as work designers: How algorithmic management influences the design of jobs. Human Resource Management

[54] Parker SK, Grote G (2022) Automation, algorithms, and beyond: Why work design matters more than ever in a digital world. Applied Psychology 71(4):1171–1204

[55] Plato (1989) Phaedrus. Flammarion

[56] Prentice R (2004) Teaching ethics, heuristics, and biases. Journal of Business Ethics Education 1(1):55–72

[57] Rességuier A, Rodrigues R (2020) Ai ethics should not remain toothless! a call to bring back the teeth of ethics. Big Data & Society 7(2). https://doi.org/10.1177/2053951720942541

[58] Sheridan TB, Verplank WL, Brooks T (1978) Human/computer control of undersea teleoperators. In: NASA. Ames Res. Center The 14th Ann. Conf. on Manual Control

[59] Shields M (1999) Long working hours and health. Health Rep 11(2):33–48

[60] Siau K, Wang W (2020) Artificial intelligence (ai) ethics: ethics of ai and ethical ai. Journal of Database Management (JDM) 31(2):74–87

[61] Stiegler B (2012) Relational ecology and the digital pharmakon. Culture machine 13

[62] Taha HA (2007) Operations research. Pearson

[63] Uhde A, Schlicker N, Wallach DP, et al (2020) Fairness and decision-making in collaborative shift scheduling systems. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp 1–13

[64] Vagia M, Transeth AA, Fjerdingen SA (2016) A literature review on the levels of automation during the years. what are the different taxonomies that have been proposed? Applied Ergonomics 53:190–202. https://doi.org/https://doi.org/10.1016/j.apergo.2015.09.013

[65] Vakkuri V, Kemell K, Kultanen J, et al (2019) Ethically aligned design of autonomous systems: Industry viewpoint and an empirical study. CoRR abs/1906.07946

[66] Viitasalo K, Kuosma E, Laitinen J, et al (2008) Effects of shift rotation and the flexibility of a shift system on daytime alertness and cardiovascular risk factors. Scandinavian journal of work, environment & health pp 198–205

[67] Wenstøp F (2010) Operations research and ethics: development trends 1966–2009. International Transactions in Operational Research 17(4):413–426

[68] Wolbeck LA (2019) Fairness aspects in personnel scheduling. Discussion Papers 2019/16, Free University Berlin, School of Business & Economics

[69] Yu H, Shen Z, Miao C, et al (2018) Building ethics into artificial intelligence. arXiv preprint arXiv:181202953