Mollen, Anne; Hondrich, Lukas

Working Paper

From risk mitigation to employee action along the machine learning pipeline: A paradigm shift in European regulatory perspectives on automated decision-making systems in the workplace

Working Paper Forschungsförderung, No. 278

Provided in Cooperation with:
The Hans Böckler Foundation

Suggested Citation: Mollen, Anne; Hondrich, Lukas (2023) : From risk mitigation to employee action along the machine learning pipeline: A paradigm shift in European regulatory perspectives on automated decision-making systems in the workplace, Working Paper Forschungsförderung, No. 278, Hans-Böckler-Stiftung, Düsseldorf

This Version is available at:
https://hdl.handle.net/10419/271012

Hans Böckler Stiftung
Mitbestimmung · Forschung · Stipendien

WORKING PAPER FORSCHUNGSFÖRDERUNG

Number 278, March 2023

From risk mitigation to employee action along the Machine Learning Pipeline

A paradigm shift in European regulatory perspectives on automated decision-making systems in the workplace

Anne Mollen and Lukas Hondrich

This Working Paper at a Glance

Automated decision-making (ADM) systems in the workplace aggravate the power imbalance between employees and employers by making potentially crucial decisions about employees. Current approaches focus on risk mitigation to safeguard employee interests. While limiting risks remains important, employee representatives should be able to include their interests in the decision-making of ADM systems. We are introducing the concept of the Machine Learning Pipeline to demonstrate how these interests can be implemented in practice and point to necessary structural transformations.

FORSCHUNGS-
FÖRDERUNG

Georg-Glock-Straße 18, 40474 Düsseldorf, Germany
www.boeckler.de

ISSN 2509-2359

Content

Summary

Increasingly, so-called “People Analytics systems” are being introduced into workplaces, thereby subjecting employees to different forms of algorithmic management. Such systems are used for automated decision-making (ADM) and are often based on Machine Learning (ML) algorithms. They are used to automatically scan CVs during the hiring process, allocate shifts to employees, conduct work performance evaluations, select employees for educational programmes or promotions and they might even be used to decide who to lay off.

Such systems promise to produce insights based on the processing of large employee-focused datasets, to bring efficiency gains and to make decisions on employees more objective and fact-based. That seems to be why employers and human resource departments view such systems as valuable assistants to help manage the workforce. However, the decisions these systems take in the workplace impact employees and their professional development, while in most cases they do not allow for oversight into how decisions are reached. That is why using ADM systems in the workplace comes with several serious risks for employees – a fact that European legislators, among others, have recognised.

In current Artificial Intelligence (AI) focused legislative proposals, such as the European Union’s Artificial Intelligence Act (AI Act), ADM systems in the workplace are considered to pose high risks to individuals and, therefore, require legislation to safeguard workers. Such safeguards are essential, especially since ADM systems lever out traditional forms of employee representation and co-determination. Due to the opacity and lack of oversight of these systems, they further advance the power imbalance between employers and employees as employees are the ones being subjected to the decisions of these systems.

From an employee perspective, a risk mitigation approach can only be considered as preventing the most severe harm. Instead, employees and their representatives should also be able to incorporate their interests into the decision-making of People Analytics systems. This means employee representatives need to be able to voice and integrate their interests when ADM systems are developed for their workplace.

The opacity of ML systems, especially, needs to be discussed. Sometimes not even the developers of these systems know what criteria their systems use when making decisions. While these are valid concerns, they do not mean that there cannot be any oversight or transparency, and they should not rule out using a co-participatory approach when designing these systems. In this paper, we are using the concept of the ML Pipeline

to demonstrate how employees and their representatives can incorporate their interests into specific ADM systems.

The ML Pipeline differentiates the following five sequential steps that are taken during the lifecycle of common data-driven ADM systems: (1) Problem definition, (2) Data, (3) Model Training, (4) Deployment, (5) Retraining. When specific ADM systems are being planned and developed, employees and their representatives should develop employee-focused positions and demands for each of these steps in order to incorporate their interests into the systems under development.

Within the "problem definition" phase, the precise task and objectives of an ADM system are being decided on. Here, employees should have a say in whether the problem in question is suitable for ADM and what its purpose is and they should also be able to discuss the general approach and methodology to be applied. Employees should also be aware of long-term organisational power shifts, when knowledge on organisational decision-making (for instance, who to promote) becomes increasingly internalised into an ADM system and is no longer available and transparent to employees.

The "data" stage covers questions on data collection, the operationalisation of key constructs for the decision-making of the system and data processing. At this stage, very important questions regarding the privacy of employees and the extent of surveillance they are going to be exposed to will be discussed. While it will certainly be in their interest to limit the amount of workplace surveillance, employee representatives must be involved at this stage. This is because, for some ADM systems, it might make sense to collect sensitive information on employees to later use to prevent discrimination. It is clear that such difficult decisions, and trade-offs, have to be taken on a case-by-case basis with the involvement of the people affected.

During "model training", the mathematical function that best serves the purpose of the system, as defined in the problem definition phase, is extracted. Such functions can be more or less opaque and complex. Given that useful patterns in ML systems might get entangled with harmful patterns, employees need to ensure that the extracted function does not counter their interests – for instance, by inheriting pre-existing biases from past datasets.

In the "deployment" and "retraining" phases, the system is put into practice. Here, safeguards need to be taken against the deterioration of system performance, and it is essential that the people affected can scrutinise the system to avoid the spread of harmful patterns. In case of significant system performance drops, retraining the system can be considered a

useful countermeasure – although this is a decision that should be made with the involvement of employees.

Considering how employees and their representatives can articulate and implement their interests along the ML Pipeline when confronted with ADM systems in the workplace requires a shift in the current discussions on AI legislation and on how to address risks associated with algorithmic management.

While current legislation puts forward a risk mitigation approach, trade union discussions focus on addressing the ethical standards of workplace ADM systems. While such discussions have proven insightful, a step toward more practice- and process-oriented approaches is urgently needed. These approaches need to be substantiated by further legislative attempts to empower employees in the face of algorithmic decision-making. For instance, this can be achieved with encompassing transparency requirements, support for co-participatory governance approaches, or funding for ADM-focused training programmes for employee representatives.

1 Automation in workforce management: Moving beyond risk mitigation

Managing employees is complex, requiring complicated distributional or evaluative responsibilities. These include shift and task allocation, promotion and training decisions, work performance evaluations, contract termination, etc. With such complex responsibilities to tackle, and in the hope of efficiency gains, companies are increasingly looking at using rule-based or Machine Learning (ML) algorithms to assist in workforce management.

In response, trade unions across the globe have started to consider the effects of such algorithmic management approaches (AlgorithmWatch 2023). Their main concerns are an increasing power imbalance between employees and employers, due to the opacity of algorithmic management systems, the datasets such systems rely on, and the increasing amount of employee surveillance needed to create them.

While the entities that provide ADM systems set the parameters for decisions about employees that are potentially far-reaching, the decision-making process is often opaque to those affected. In 2022, one in three employees in Europe did not know whether an algorithmic management system was being used on them or not (Holubová 2022). To remain effective when being confronted with ADM systems, implementing employee interests through co-determination, social dialogue or collective bargaining need to be adjusted.

ADM systems in the workplace can have severe negative impacts on employees – from discrimination to faulty decisions without effective redress mechanisms to opaque decision-making without sufficient accountability instruments in place. Across trade unions, there seems to be consensus about a number of guidelines and ethical principles that should orient employee representatives when they are confronted with algorithmic management systems and the risks they pose. Roughly summarised many unions seem to agree on transparency, fairness, privacy, data protection, the right to be informed, human oversight, high-risk regulations, explainability, etc. as safeguards to implement when ADM systems are introduced in a workplace (AlgorithmWatch 2023). These principles are, first and foremost, a risk mitigation strategy.

Also, current legislative proposals, including the European AI Act, mostly focus on minimising the risks associated with ADM systems. At present, the proposed European regulation foresees a self-assessment process to ensure the alignment of high-risk AI systems with fundamental

rights. To a large extent, AI systems used in the workplace are considered to pose high risks, especially when it comes to recruitment, advertising vacancies, screening or filtering applications, evaluating candidates, for promotion and termination matters, for allocating tasks, and for monitoring and evaluating the performance and behaviour of employees (European Commission 2021a). While it is important to note that the AI Act recognises the risks of ADM systems in the workplace, it only provides the most basic safeguards.

Employee representatives should be able to move beyond a risk mitigation approach and identify potentials for intervention in the interests of employees. Their involvement should aim to shape ADM systems in the interests of the people affected. However, current legislative proposals on AI, such as the European AI Act, are inadequate and further European or national directives are needed.

That is why we argue for a paradigm shift in European regulatory perspectives on ADM systems in the workplace. Mitigating and limiting risks are essential safeguards but we need to move forward – by enabling employees and including their interests whenever ADM systems are designed for the workplace. The urgent question is thus, how can employees and their representatives actively articulate and include their interests in an ADM system, specifically when managing the workforce? Or to speak more precisely: Where is a space for action on behalf of employee interests when it comes to ADM systems used for so-called People Analytics procedures?

This is not a trivial question. ADM systems are often described as black boxes that allow little insight into how they operate and even fewer options for shaping their outcomes. ML systems learn autonomously, often in an unsupervised manner. This means they could, potentially, constantly change the logic they use to operate. As a result, there are concerns that advocating for employee interests, with regard to such systems, could become impossible. However, these assumptions are based on a number of myths surrounding ADM systems that require debunking to overcome apathy and a feeling of powerlessness when it comes to advocating for employee rights in relation to ADM systems.

In the next section of this paper – following a brief definition of key terms – we will identify where ADM systems open up spaces to advocate for employee interests. To do this, we will use the concept of the ML Pipeline to provide transparency and break down ADM systems into workflows and stages within an overall lifecycle, where employees can potentially include their interests.

Our goal is to demonstrate how employee representatives can find specific entry points on a technical and organisational level to advocate

for their interests. At the same time, this implies that employee representatives are capable of articulating precise interests that can be implemented. We will underline these reflections with fictional, but realistic, workplace examples to illustrate how the ML Pipeline can be used to provide transparency and articulate employee interests. We, then, provide an outlook on how more participation from employees and the inclusion of their voices can be achieved – and how this relates to provider and user obligations in AI regulations.

2 People Analytics: Risks and transparency for affected people

New technological developments often tend to be connected with an air of omnipotence and almost unprecedented magical capabilities – which holds especially true in the case of so-called Artificial Intelligence (AI) (Campolo/Crawford 2020) or ADM systems. In the case of ADM systems, this mystification of technology is problematic in two ways. First, it promotes overestimating the capabilities of ADM systems. Second, it prevents people from gaining a clear understanding of what ADM systems do and how they work. Therefore, it is paramount to define precisely what we are talking about when we address automation in the workplace. We need to have a very clear understanding of our objects of inquiry: ADM systems and People Analytics procedures.

2.1 ADM systems

To avoid the pitfalls created by mystifying the term AI, we will speak about ADM systems. We understand the automation part to entail either rule-based algorithms or ML systems – this also includes Neural Networks or Deep Learning systems. Even though many problems also concern (more) simple rule-based algorithms, in this paper we will focus on ADM systems based on ML.

In short, “Machine learning is an approach to learn complex patterns from existing data and use these patterns to make predictions on unseen data” (Huyen 2022). These patterns might capture simple correlations, e. g. between the years of work experience and compensation or more complex, non-linear relationships, e. g. between a person’s work experience, skills or work performance assessments and their suitability for vacant jobs. Generally speaking, these systems tend to be more complex when it comes to transparency, oversight and participation.

2.2 People Analytics software

The basis of this paper is grounded in our work on the use of ADM systems in the workplace, specifically for managing the workforce algorithmically. We refer to such systems as People Analytics systems, but other terms – Human Resource Analytics, Workforce Analytics, etc. – are commonly used. All these terms describe software-based systems that use

data analysis and automation to exercise typical human resource-related tasks to manage the workforce.

These systems process workforce-related data to generate and visualise insights and information about the workforce – this can encompass descriptive, predictive, and prescriptive elements. Based on the collected and analysed data, these systems can also make assumptions and decisions about the workforce (Gießler 2021). Therefore, to use a People Analytics system, relevant workforce data have to be monitored and collected.

The use of People Analytics systems can range from using automation during the recruitment process – through the algorithmic scanning and sorting of applications – to automating shift plans, allocating tasks, evaluating productivity and achievements, preventing work-related accidents, proposing further training or promotions or predicting an employee's loyalty to their employer and their likelihood of quitting their job, etc. The promises that providers of People Analytics systems put forward are quite far-reaching, as are the ways in which they are applied, some of which can have a potentially severe impact on organisational structures and the power dynamics between employees and employers (Jarrahi et al. 2021).

2.3 Risks in People Analytics

In its current form, the AI Act recognizes that risks are associated with ADM systems in the workplace. The legislative proposal specifically sees high risks in the following areas:

> "(a) AI systems intended to be used for recruitment or selection of natural persons, notably for advertising vacancies, screening or filtering applications, evaluating candidates in the course of interviews or tests;
> (b) AI intended to be used for making decisions on promotion and termination of work-related contractual relationships, for task allocation and for monitoring and evaluating performance and behaviour of persons in such relationships." (European Commission 2021a, p. 4)

By focusing specifically on any use of ADM systems that might have an impact on a person's forthcoming, this list touches upon some of the major risks associated with ADM systems. This includes risks associated with education and training, looking for employment, being employed and being subjected to task allocations and job performance evaluations, and also covering job termination.

The idea of the Act is that providers of such high-risk AI technologies need to comply with certain requirements in order to limit the potential risks associated with their use. Such requirements potentially pertain to

data and data governance (article 10), technical documentation (article 11), record-keeping (article 12), transparency and provision of information to users (article 13), human oversight (article 14) as well as accuracy, robustness and cybersecurity (article 15) (European Commission 2021b).

A major shortcoming of the AI Act's risk-based approach lies in its focus on bringing products onto the market. AI systems need to comply with certain requirements in order to be made available in the European Union. This puts responsibility predominantly on the providers of AI systems and not on the users – in the language of the AI Act, a user would be a company using a People Analytics system to manage their workforce.

Therefore, with respect to AI systems in the workplace, the AI Act is about attempting to minimise risks and not about empowering employees to articulate and include their interests whenever People Analytics systems are used. But risk mitigation cannot suffice. It is the lack of employee voices when implementing work-related ADM systems that needs to be addressed.

2.4 From mitigation to voice

When it comes to high-risk systems, it is hard to imagine that the requirements in the AI Act regarding data, data management and data collection (article 10) are sufficient from an employee perspective – especially considering the various workplace surveillance practices already in place. The vast extent to which data is monitored and collected should raise major concerns about the level of workplace surveillance and the degree to which deductions, analyses and profiles are automatically generated through People Analytics systems.

A 2021 report by Cracked Labs, a non-profit organization, gives a comprehensive and worrisome overview of the People Analytics systems on the market and in use (Christl 2021). The report lists surveillance practices that are necessary to operate specific ADM systems in the workplace. These range from monitoring devices employees carry on their bodies, workspace cameras that record the movement of employees, systems that monitor mouse movements and clicks, scanning of communication channels, including non-work-related social media activities, the extended use of log files, etc. (see also Krzywdzinski et al. 2022). Not all of these far-reaching surveillance practices likely comply with current or proposed European legal frameworks. But even if they do, their risks should still be addressed.

Alongside the pervasive collection of employee data come risks associated with linking different employee-related datasets across systems

and platforms, resulting in extensive profiling of employees and complex analyses of their data. In this regard, the lack of transparency, accountability and participation from employee representatives is worrisome. It is often unclear how People Analytics systems operate and take decisions, whether the decisions made are solid, fair, and justified, or if employees are aware of them and if there are any safeguards to counteract problematic or incorrect decisions.

The AI Act foresees the making of a risk assessment before a product arrives on the market. However, this is not a sufficient safeguard. ADM systems need to be examined on a case-by-case basis while systems are in use. Strengthening existing employee representative structures to voice the concerns of employees when it comes to workplace ADM systems seems like a promising direction. However, previous engagement with employee representatives and labour organisations shows that the lack of transparency associated with these systems, and insecurities about how the interests of employees can be incorporated into them, make it difficult for employee representatives to voice their apprehensions.

2.5 Transparency as a precondition

The topic of transparency and how to make information on ADM systems available for employees and their representatives figured prominently in the initial European White Paper on Artificial Intelligence (European Commission 2020a). The topic remains a major requirement for high-risk systems in the AI Act – as a means to minimise potential harms associated with these systems (see, among others, article 13). Beyond a risk mitigation approach, it is a precondition that employees articulate their interests during the design and implementation process of People Analytics systems.

In the following section of this paper, we will use the concept of the ML Pipeline to suggest how to achieve transparency. In addition, we will look at ways in which employees and their representatives can use this transparency to voice their concerns during the design and implementation of ADM systems in the workplace. But before that, we will briefly summarize ongoing discussions by labour organizations on the subject of People Analytics systems.

3 Labour perspectives on how to implement employee interests in ADM systems

The current negotiations on the AI Act offer an opportunity to assess where labour organisations stand regarding ADM system regulation. During the negotiations on the AI Act, all interested stakeholders and individuals had the opportunity to submit statements during a public consultation process. A number of trade unions and trade union confederations and other civil society stakeholders submitted specific labour-related positions on the AI Act, which we analysed. At the same time, we systematically scanned eight European countries (Czech Republic, Estonia, Germany, Hungary, Italy, Poland, Spain and Sweden) to see how labour organisations engaged with the issue of algorithmic transparency and accountability in the workplace, be it in relation to the AI Act or not.

We aimed to condense the positions of the relevant labour stakeholders on legislative approaches to address employee safeguards regarding the implementation of People Analytics systems. In addition, we addressed how employees can articulate their interests in practice regarding the design and deployment of ADM systems beyond a risk mitigation approach. We specifically engaged with stakeholders who have experience in the domain both at conceptual and shop levels. We intended to move from a theoretical to a practical level, generating ideas on concrete instruments and processes. For instance, how to put meaningful ADM system transparency into practice and how employee representatives can meaningfully advocate for their interests when introducing People Analytics systems into the workplace.

Based on our engagement with these labour stakeholders we conceptualized how ADM systems along the ML Pipeline can empower employees and their representatives in articulating their interests. In the following section of this paper, we will summarise our major insights and deductions.

3.1 Insight 1: Common but often vague risks and concerns

Many stakeholders identify similar risks related to ADM systems in the workplace. Such apprehensions are reflected in the public consultations submitted as part of the negotiations on the AI Act (European Commission

2020b) and in related publications by trade unions such as industriAll (2021), UNI Europa (2019), the German Trade Union Confederation (2021), European Trade Union Institute (Ponce del Castillo 2021), etc. These concerns are also reflected in the current activities of labour organisations on algorithms used in the workplace that we systematically analysed across eight European countries (AlgorithmWatch 2023).

Common fears concern job loss and deskilling, due to increasing workplace automation, the dangers of discrimination and false decisions made by ADM systems or fears about the increase in workplace surveillance and violations of privacy. At a more abstract level, there are more general concerns about ML systems, specifically the danger that such systems might learn in unsupervised ways, leading to a loss of control and lack of insight for both the people deploying the systems and those being targeted by them. The discussion seems homogenous, but the risks discussed remain quite general.

3.2 Insight 2: Focus on risk mitigation

With the risk-based approach of the AI Act, it is not surprising that the public consultations commenting on the draft proposal focus on how to mitigate risks associated with workplace AI systems. For example, several trade unions have proposed transparency requirements, auditing procedures for high-risk systems, and privacy protection measures (e. g. prohibition of employee tracking, biometric technologies, and wearables) to protect employees from the risks mentioned above (see for instance, Association of Nordic Engineers 2021, German Trade Union Confederation 2021).

Beyond negotiating the AI Act, further statements and publications by labour organisations mostly discuss safeguards regarding the risks associated with automation in the workplace. Such risk mitigation strategies are, for instance, proposed as ethical or regulatory principles that are supposed to construct frameworks along which ADM systems in the workplace should be developed and implemented (AlgorithmWatch 2023). Their focus lies mostly in mitigating the risks of ADM systems in the workplace, for instance, by guaranteeing human oversight or privacy protection and by ensuring transparency, autonomy, fairness, and security.

3.3 Insight 3: First calls for participatory approaches

So far, what is not big on the agenda are calls for participatory approaches that would allow employees and their representatives to create safeguards against the risks and dangers of ADM systems in the workplace and allow employees to shape them according to their interests. For example, during public consultations about the AI Act, the German Trade Union Confederation (2021) and the Association of Nordic Engineers (2021) called for participatory governance approaches. Beyond the AI Act, the Swiss Trade Union Syndicom (2020) highlighted social partnership and employee participation as one of nine guiding ethical principles for the development and deployment of ADM systems in the workplace.

Such calls are underpinned by references to established co-determination practices and the need for employees to be able to make their voices heard, especially where ADM systems touch upon fundamental ethical questions. However, despite calls for encompassing transparency, there are still very few suggestions on how to enable such participatory governance approaches. The Trade Union Confederation (TUC 2022) in the UK is one of the first labour organisations to have issued a hands-on action guide for negotiators on collective agreements and ADM systems.

In order to move away from general fears and mitigation strategies, more such practical guidelines are needed. In the following, we will use the ML Pipeline concept to demonstrate how feasible participatory governance approaches to People Analytics systems are.

4 Employee action along the ML Pipeline

It is of utmost importance to establish safeguards to mitigate the risks associated with People Analytics systems. In the workplace, it is essential to go one step further – by addressing how employee representatives can advocate for employees' interests when they are confronted with these systems. There are potential benefits associated with ADM systems and employees should be able to make use of such benefits. This implies that employee representatives can translate and implement their interests into People Analytics systems. In the following part of this paper, we use the concept of the ML Pipeline to show how employees can find entry points for co-participatory approaches in shaping specific People Analytics systems.

The ML Pipeline is a concept that can be used to increase the transparency, auditability, and governability of ML systems (for a discussion on bias, see Schelter/Stoyanovich 2020). It outlines five sequential steps that are taken during the lifecycle of common data-driven ADM systems:

- Problem definition
- Data
- Model training
- Deployment
- Retraining

These steps are defined in a generic way to apply to a wide range of data-driven ADM systems. This differentiation can then be used to analyse the implications of an ADM system and the impact on the people it is being used on. Furthermore, these steps can be used to design ML systems so that they serve the interests and fundamental rights of all parties involved.

Figure 1: Employee interests along the Machine Learning Pipeline

/ ML-PIPELINE	/ EXAMPLES FOR EMPLOYEE INTERESTS ALONG A ML LIFECYCLE
1 PROBLEM DEFINITION	What problem should be solved and how?
2 DATA	What data are used and how are these data to be interpreted?
3 MODEL TRAINING	What are the best methods for the system to achieve the defined goal?
4 DEPLOYMENT	How are the results of the system used in operations?
5 RETRAINING	What quality assurance standards will be in place?

Source: Own illustration

At the moment, the usual purpose of ML systems used for People Analytics procedures is to advance business interests, e. g. to reduce costs or expand capabilities. It is important to note that every step taken throughout the lifecycle of such an ML system is somehow interlinked and geared toward this objective. In parallel, from the definition of a business case for an ADM system, to choosing the right methods, operationalizing key concepts, gathering data, training the model, taking action, measuring outcomes and maintaining the system, the interests of employees must be considered. That is why the interests of employees cannot be sufficiently implemented through selective interventions.

Instead, the ML Pipeline can assist in completely thinking through and designing an ML system, with all its stages and workflows, from an employee perspective. In doing so, employee representatives might succeed in advancing employee participation whenever workplace ADM systems are planned, developed and used.

4.1 Problem definition

The purposes of the problem definition phase are to define the problem space, deduct tangible objectives and specify how an ADM system can be designed and integrated into an organisation to advance the identified objectives. These discussions usually include negotiations with stakeholders, to which affected people such as employees should belong. In addition, the computational and personnel resources needed to accomplish these objectives are defined. Also, specific methodology and technical re-

quirements (e. g. minimal performance conditions) and the scope of the system are discussed.

Due to the explorative nature of the development of ML systems, some projects define these parameters vaguely or condition them on intermediate results. Therefore, during the entire development process, flexibility and the ability to react to unforeseen problems remain essential. For employee representatives, it is especially important that they voice their interests during the problem definition phase. It is at this stage that fundamental questions regarding the purpose and scope of ADM systems and subsequent organisational changes are negotiated.

4.1.1 Setting the ground

Because of its non-technical nature, the problem definition phase is often not regarded as part of the ML Pipeline. However, since this is the main planning stage where key directions and approaches are defined, we deem it a critical stage for voicing stakeholder interests. Defining a problem and finding a workable technical and organisational solution is a potentially highly contested moment during negotiations. Power imbalances, lack of voices, and inadequate representations at this stage will persist throughout the entire lifecycle of an ADM system.

For example, the integration of many diverse stakeholders, including employees or historically marginalised groups, might be crucial in order to avoid discrimination through ADM systems. When such voices are integrated into the planning process, attention to possible discriminatory results might be heightened and adequate fairness metrics could then be integrated into the system.

4.1.2 Long-term power shifts

The introduction of ADM systems in an organisation not only means automating processes, it often implies the re-structuring of organisational processes. This can mean long-term ramifications for power relations within the organisation, that employee representatives should keep an eye on.

When an organisation introduces an ADM system to automate parts of the internal and external hiring process, employees might lose valuable knowledge of hiring procedures. When a single system manages the internal job market – instead of hundreds of people spread around the company – it could result in negative consequences for the negotiating power

of employees. While the ADM system, and the people working with it, might increasingly centralise relevant data and insights about careers and the potential of employees, the information on the supposed worth of those employees is kept from the staff themselves. Such information inequalities might affect the ability to reach fair compensation rates.

Therefore, the problem definition phase is not only about defining the scope of one individual ADM system, it means getting all stakeholders involved in a discussion about how new ADM systems might among other things change the flow of information and decision-making procedures in a workspace, possibly resulting in long-term organisational changes and power shifts.

4.1.3 Scope creep

Initially, the intended use or application of a project may be limited in comparison to what it eventually becomes used for. For example, a system designed to predict outcomes might later be used to prescribe action. This is problematic as it opens the door for correlations being mis-used as causal predictors – which potentially can lead to discriminatory decision-making. If we think about correlations in the workplace, we for instance often see a correlation between gender and a specific field of work. For example, there is a high probability that the proportion of male employees in an IT department is higher than in other departments.

However, it is inadmissible to infer a causal relationship from such a correlation. If an ADM system designed to select job applicants were to understand this correlation as a causal relationship, it might favour male applicants when selecting personnel in the future. Notably, this may also happen when gender is not an explicit input-variable – for instance when variables correlated to gender are available (e. g. some names of schools are gendered, hobbies might be correlated to gender). Therefore, the risk of making false assumptions about what ML models can be used for is very high. Correlations might be sufficient to understand and predict outcomes. However, to prescribe actions, a causal model is needed (Barocas et al. 2019).

Thus, it is critical to be realistic when planning data-driven projects and to keep in mind what an ADM system is designed to achieve and to clarify and monitor very carefully what its areas of application are.

4.2 Data

The purpose of the data collection phase is to prepare a dataset that maximises the amount of useful information needed to train the ML model in light of the objective of the ADM system. Before using training data, the data must be collected, pre-processed and transformed so that it can serve the purpose of the ADM system.

It is important to critically assess how the data in question has been generated. Consulting employees who generated the data might be a crucial step. They have the relevant expertise and understand the context to help interpret the data. For instance, employers might feel tempted to use already existing datasets of employee communication (e. g. mail, workplace chat) to draw conclusions on communication and performance of employees, however, this is problematic for many reasons, i. e. it might constitute a privacy violation, it might be easily gamed and it might introduce biases, e. g. against people using other means of communication.

All decisions on data selection and collection and privacy protection are highly significant for employees, the surveillance they might encounter in their workplace, and the output of the ML system. For ADM systems to benefit employees, there must be adequate oversight. This includes allowing employee representatives to have their say over the nature, and quality, of the data selected. Documentation of datasets in the form of so-called data sheets or data cards (Gebru et al. 2021) is an attempt to increase the transparency and safety of their use and might assist all the stakeholders involved.

4.2.1 Operationalising key constructs

During the design and data management process, it is necessary to decide what key constructs shall be used for the ADM of the system in question. The basis for this decision is, of course, the objective of the system as defined in the initial phase of the ML Pipeline. But it also depends heavily on the data available, the data that can be collected in the future and the processing possibilities that these datasets offer. To advance their interests, employees and their representatives should think about how their goals can be operationalised. For example, in the workplace, employee representatives might like to implement a tangible metric that best captures their preferred notion(s) of fairness in an ADM system

In other cases, employees might be interested in systems that are geared toward improving work culture. In such a scenario, employees are the experts, and they need to be consulted as the ones being affected by

the decisions taken by the ADM system. They will have specific ideas of how to operationalise key constructs – for instance how best to measure "work performance", "productivity", "suitability" for vacant positions, etc. They could rely on established psychological constructs like conscientiousness, openness or intelligence that have been operationalised in peer-reviewed, benchmarked behavioural tests or questionnaires. However, pseudo-scientific theoretical constructs like phrenology – falsely assigning potential personality traits to facial features – would be unacceptable.

It is important to note that, even in cases where unsubstantiated theoretical constructs are not explicitly stated and modelled, the data and ML model might still internally compute and rely upon them. Therefore, it is necessary to critically ask for the justification for the gathering of specific data sources, i.e., photos or names of employees or applicants. Returning to the idea that it is possible to measure productivity by analysis of work messages – this method of collecting data probably favours people who work more remotely, because they produce more data traces than people sitting opposite each other in an office.

These unintended consequences need to be considered. To meaningfully interpret and contextualise existing data traces often requires knowledge of the domain. That is why, employees as stakeholders should be included in establishing an informed theoretical understanding of how ADM systems define, operationalise and measure key concepts in their decision-making.

4.2.2 Data collection

Data is the foundation of the patterns extracted by ML models and, thus, the output generated by ADM systems. Therefore, data collection and processing are essential for building solid ADM systems. As can be seen in many examples, when historical bias seeps into training data, it can lead to the perpetuation and reinforcement of those biases. Employee representatives can advocate for the collection of data on sensitive subjects like gender, race and age in order to detect and then reduce these biases. Collecting employee data on sensitive or even protected attributes can be critical, because it might make harms visible and their objectives optimisable. However, this may cause difficult trade-offs between these objectives and privacy or data protection laws – demonstrating again the need for the people affected to be able to articulate their position on a case-by-case basis.

Data can also be an important tool in promoting specific employee interests beyond preventing harm. For instance, if an ADM system is intended to suggest training courses for employees, then data about employees' interest in certain topics might need to be gathered and considered so that the system can select the most appropriate training course. That is why employee representatives should be involved in decisions on what employee data should be collected in order to assess how an ADM system might serve the people affected.

4.2.3 Data processing

Collected data for training an ADM system is usually unstructured, messy, and incomplete. To make the data usable, it must be cleansed, processed, and possibly re-structured into more complex features. The data collected might, for instance, be too granular. If an ADM system is supposed to evaluate employee productivity, then it might collect keyboard strokes, messages sent between employees, and the time stamps of this information. But the time stamp as such might not be very informative, so data might be aggregated to differentiate between different periods of the working day or week.

Multiple methods and procedures exist for processing data and many of those processes involve decisions that can for instance significantly impact both the performance and fairness of the ADM system, thus touching upon the interests of employees. For instance, within a dataset, it is common for some values to be missing. This could be due to problems collecting data, but it could also be because sensitive information, e. g. regarding ethnicity or sexuality, is deliberately omitted from a dataset. Such missing values might have negative effects on the performance of an ADM system.

Missing values in data are often replaced with modelled data through imputation methods. However, this process can introduce bias towards groups underrepresented in the dataset (Martínez-Plumed 2021). To promote fairness in terms of sensitive attributes, counterfactual samples can be added to the dataset to address underrepresentation of specific demographic groups and reduce bias. These methods however only work imperfectly and applying them correctly is far from trivial.

At this stage, it is clear that advocating for employee interests requires deep technical knowledge of specific ML and data management procedures. This points to the need for employee representatives to consult external experts and the urgency for some to be trained more thoroughly in the basics of ML methods.

4.3 Model training

The purpose of the model training phase is to extract the rules, statistical patterns, and contingencies to optimise the defined objectives of an ADM system. The result of this process, the Machine-Learned model, is a mathematical function that might be more or less complex and opaque, depending on the algorithm used. Important sub-components here are the training algorithm used to steer the process of optimising the model, the objective function (or loss function), defining the goal of this optimisation process and the metrics helping the engineers get a clearer picture of different aspects of the optimisation process.

The general problem here is that useful patterns in Machine-Learned systems might be entangled with harmful patterns. Furthermore, whether a pattern is harmful or useful is also highly context-dependent. With current methods, disentangling useful and harmful patterns is only possible with limitations. Also, in many cases, trade-offs between legibility and performance might exist. However, even if more complex and opaque models are used, resources can be invested to render input-output contingencies more transparent, explainable, and robust.

Employees can promote their interest in two ways: by preventing harmful biases from seeping into the Machine-Learned model, and by specifying a model training procedure geared toward picking up useful data patterns geared towards their objectives. While ML is a vast and fast-moving field with a great diversity of systems, for most systems it is possible to single out three components that drive the later functioning: objective function, optimization algorithm and metrics.

4.3.1 Objective function

The objective function is a mathematical representation of the goals of the system. During the development of an ADM system, employee representatives should ensure that this mathematical representation does not perpetuate harmful biases. Furthermore, it should also promote their specific interests beyond safety, e. g. by considering and optimizing for their personal interests and career goals.

To use the example of an internal hiring system again, the approach of an ADM system could be to perform a semantic search between the requirements of an open position as well as the documented qualifications of employees. This process encodes the skills that employees may have and the qualifications that jobs may require in a common semantic space. In a second step, matches between employees and vacant positions

could be generated by identifying skills and qualifications that are as close together as possible.

In order to do so, the text input data of the entities to match (e. g. documentation about employee skills and job descriptions) need to be transferred into a common semantic embedding space, while capturing the semantic meaning of the text data. After encoding the text data in the embedding, pairing can be generated between text descriptions of employee skills and skill requirements in job descriptions to minimise the distance between the pairs – the closer the distance, the more semantically similar the job and skill descriptions are.

A common way to obtain such an embedding space is by trying to predict words based on other words in their respective neighbourhood – the objective function would then be the maximisation of the likelihood of this prediction. The problem here is that the prediction is based on training data and, thus, it learns to perpetuate biases present in this data. However, there are ways to mitigate such a bias.

In the example, it would be possible to add another objective. The goal of the system should not only be to maximise the likelihood of predicting words in each other's neighbourhood. But another objective could be to minimise the distance between, for instance, gendered words to one another and equalising the distance of gendered words to gender-neutral, normative words in the embedding space. This would result in reducing bias based on gender. Also, it is possible to expand the objective function to make the matching process fairer, e. g. by minimising the correlation between group membership (for instance, by gender) and the outcome of the recommendation (cf. Bolukbasi et al. 2016, Caton et al. 2020).

Now, to promote the interests of employees the objective function could be further expanded to also optimize career goals of employees. This could be analogously done by performing a semantic search between interests and goals of employees and affordances of the vacant positions.

In theory there is no limit of how many objectives can be optimized, in practice however, trade-offs will have to be made. Employees and their representatives should be aware that the objective function might be a crucial target to advocate for their interests.

4.3.2 Optimisation algorithm

Objective function and the optimisation algorithm are closely related concepts. However, while the objective function represents the target, the optimisation algorithm defines the solution space and the path a model "takes" through the solution space towards the objective during optimisa-

tion. Usually the optimisation is imperfect, and so the training algorithms heavily influence at which point the model ends up in the optimisation procedure. This has ramifications on the robustness, fairness and observability of the resulting model.

For instance, bigger models, like large language models, might be more capable and expressive than smaller ones, but harder to test and to debias. Adapting and finetuning them might be computationally expensive. Not only the model size, in other words the number of parameters the model consists of, but also the architecture can matter for some purposes: When generating counterfactual examples to increase fairness, for example, methods might specifically differ in the fidelity with which they can model and generate counterfactual examples in intersections of demographic dimensions (e. g. race, sex and age) (Creager et al. 2019).

At this stage, it is probably again necessary for employee representatives to get support from ML experts on the specifics on training methods and what purpose they potentially serve.

4.3.3 Metrics

Metrics should be used to track and guide the experimental and iterative modelling process to stay close to the high-level objectives formulated by the stakeholders and may especially cover aspects that might not already be covered by the objective function. Metrics should capture the whole breadth of the problem space and facilitate interpretability of the optimisation process – including aspects that come from the employee side – to minimise the chance of unwanted, undetected side effects. Employee representatives should be able to have their say on the metrics used to evaluate system performance. However, this is hard to attain as it might not be possible to observe every aspect of the problem space.

For instance, in our example, it might be hard to define a successful match in an automated internal hiring procedure: Is it possible to simply ask the employee or supervisor if the employee is a good fit? Is it sufficient if an employee did not quit or got fired? How long should one wait before evaluating these things? Some aspects are intrinsically hard to assess, like the quality of rejections: There is simply no data available on the matches that did not manifest.

Still, other important data points might be readily available, for instance, who the beneficiaries of our recommendations were, and if the aggregated results satisfy the requirements that we posed concerning the fairness of these positive outcomes.

Notably, it is often a question of resources, if data is made available for a more thorough analysis of negative outcomes. Such measures of obtaining information may range from conducting post-rejection interviews to hiring a random sample to obtain experimental, less-biased data on the validity of the model (Bird et al. 2016). However, it is paramount that stakeholders raise awareness of the limitations of making the optimisation process observable via the metrics.

4.4 Deployment

The purpose of the deployment phase is to materialise the objective of the ADM system. Deploying an ADM system means integrating the Machine-Learned model into a traditional software system that handles input, output, pre-and post-processing of data, user interface, and the computation on hardware. Along with the deployment of the ADM system itself, tools for logging input and output data as well as monitoring key metrics must be in place. Only at this stage, are the model and the ADM system first run against real-world data. Having been optimised for the training data, it is likely that the system will exhibit a drop in performance with regard to the metrics that try to capture the pre-defined problem space.

That is why the ADM system should be scrutinised closely to see if it continues to meet minimal performance requirements. Relevant oversight, especially by representatives of the people affected, is essential here. Also, the user interface and user experience can matter a lot for the effect the ADM system has on its environment. In these regards, it is important to adhere to best practices and regularly monitor the influence of potential changes.

For the example of using an ADM system to assist in an internal hiring procedure, this would mean that matching scores (on employee skills and job requirements) are communicated to the management and HR department, which might then act. Therefore, in our case, the last step in the decision-making process is human. It is worth noting that this again opens the door to further biases and abuses of power – for instance, when only selectively acting on the decisions taken by the ADM system. This could be prevented by formulating clear processes for human intervention and subsequently monitoring compliance. However, even with monitored, procedurally fair processes and an ML Pipeline generally following best practices, there is no guarantee that the result satisfies certain outcome-fairness criteria.

This should further substantiate the importance of monitoring the outcomes of decisions made by the ADM system. It is important to specify

what is being monitored to serve employee interests, i.e., reaching a certain degree of outcome fairness towards sensitive attributes like gender, ethnicity, and age. Groups with such sensitive attributes – and their intersections – would need to be monitored in a disaggregated way. If outcome fairness does not reach acceptable levels, the ADM system could be further calibrated, meaning that group-specific decision boundaries could be lowered or heightened. Notably, this might impair process fairness and illustrates one of the trade-offs arising when trying to design ADM systems based on high ethical standards (Kleinberg et al. 2016).

4.5 Retraining

Retraining is a form of maintenance for the ML model and the ADM system. It is done to keep the system aligned to the original objective in the face of unexpected and complex socio-technical dynamics. While the ML model represents a static snapshot of training data at the time of training, the inference data the model is run against tends to change in most environments and contexts. This leads to a growing divergence between the data the model is optimised for and the data it encounters, leading to a deterioration in performance. A common countermeasure is to retrain the model with more up-to-date data.

How fast the ML model deteriorates mostly depends on the stationarity of the application domain of the ADM system: In the case of social media apps like TikTok and Instagram, these changes might occur within minutes, in organisational contexts, the drift might happen within weeks or months and purely physical problem domains might be entirely stationary.

Even though retraining is an appropriate method of countering the deterioration of performance, it introduces new challenges to the interests of employees. These challenges stem from the various kinds of interactions of the technical system with its environment, the organisational and social system. Data used during the retraining of the Machine-Learned model will have been influenced by the model itself. Any signal or pattern the model uses to make decisions is subsequently more represented in the output data and, therefore, the training data of the next iteration, potentially leading to emergent bias (Schelter/Stoyanowitsch 2020) and feedback loops (Barocas et al. 2019).

Already on the first iteration, this may cause problems, as predictions and decisions might become self-fulfilling prophecies – what has been documented, for instance, in predictive policing (Dobbie 2016, Ensign 2017). Through many iterations, this effect can accumulate and even become problematic in less sensitive domains. In terms of recommendation

engines at work – like the one in our example – this could mean that certain groups of employees are treated preferentially.

Countering this is an open research topic and may depend on the domain and what kind of feedback effects will be expected. Some countermeasures have been proposed, e. g. facilitating detection through monitoring the shift of input and output data, by modelling the influence of the model and correcting for it (Krauth et al. 2022) or – if possible – by only retraining on data that is uninfluenced by the model itself (for an overview of practical approaches see Huyen 2022).

From an employee perspective, it is important to acknowledge the need for oversight, control, and safeguards in the retraining phase – thus, potentially after an ADM system has been implemented and becomes established in an organisational context. This calls for an oversight process that continuously involves employee representatives in monitoring existing ADM systems as well as effective redress mechanisms for the people affected by decisions taken based on possibly deteriorating ML models. Here again, it becomes obvious how far technical safeguards need to be complemented by organisational safeguards as well as by an awareness of the shortcomings, risks, and dangers of ADM systems – even when they have been designed with employee interests in mind.

5 Beyond risk mitigation: Capacity-building and participatory governance

ADM systems used for workforce management cannot be made to work in the interests of employees without ensuring the continuous involvement of employee representatives throughout the entire development, implementation and application phases of these systems. By using the concept of the ML Pipeline, we have demonstrated the manifold decisions that have to be taken throughout these processes, and where it will be essential to enable employee representatives to have a say. This is not only decisive for designing the technology in more employee-friendly ways, but it also fosters trust among employees.

Our reflections have equally demonstrated how far ethical principles guiding the design of ADM systems in the workplace can be implemented in practice. This step from principle to practice is urgently needed and requires further substantiation through good practice examples, guidelines for employee representatives, and capacity-building initiatives. Even though employee representatives do not have to become ML developers, the above reflections demonstrate how a basic understanding of ML procedures is essential in order to ask the right questions and understand the basic shortcomings of ML models on a case-by-case basis. Thus, we need further initiatives on:

- Demystifying ADM systems along the ML Pipeline
- Identifying entry points for employee advocacy along the ML Pipeline in the interest of making People Analytics systems benefit the people affected
- Investing in capacity-building measures for employee representatives to articulate employee interests along the ML Pipeline and to raise awareness about the shortcomings of ADM systems on a case-by-case basis
- Enabling employee representatives to consult external ML experts on detailed methodological questions

Here, we have proposed a process-oriented perspective on how employees can manifest their interests in the planning, development, and implementation of People Analytics systems. This perspective can also help in assessing third-party ADM systems that are bought from external providers and are ready to be implemented. Given a sufficient level of transparency (e. g. by making use of data cards, model cards, and further documentation) at least to some extent their alignment with employee interests

can be assessed before an employer decides to buy a specific ADM system. But in order to make such an assessment, it will be helpful for employee representatives to consult how their interests can be implemented at every step of the ML Pipeline. It can also assist in retrospectively scrutinising and interrogating the objective (problem definition), data management procedure, training, and retraining modalities, etc., of a People Analytics system.

Even though it is probable that the risk-based approach of the AI Act inevitably leads to risk-mitigating perspectives, it has opened up occasional calls from trade unions and labour organisations for more participatory governance approaches regarding ADM systems in the workplace. Such calls reflect the much-needed move from formulating ethical principles along which ADM systems should be applied in the workplace, to a hands-on perspective on how employee representatives can apply such ethical principles in practice.

On the one hand, it shows again how the risks and potentials of ADM systems need to be assessed on a case-by-case basis and can only be abstracted across areas of application at a very basic level. On the other hand, considering how employees can manifest their interests in People Analytics systems along the ML Pipeline, this demonstrates how far the ex-ante logic of the AI Act and other risk mitigation AI policies necessarily remain limited in scope and are eventually insufficient for truly safeguarding employee interests.

Such risk-based approaches might end up spanning a very basic safety net preventing ADM systems that would pose very obvious risks and harm to the affected employees from entering the market. This cannot be considered sufficient from an employee perspective. Instead, employee representatives and stakeholders need to think beyond risk mitigation and regulations such as the AI Act. They need to think about how they can start advocating for and implementing employee interests within People Analytics systems so that they are not only protected from their risks but might also benefit from their potential.

Further national or transnational initiatives, among others considering the respective national context of employee representations, are necessary, and they should be designed to mitigate risks and empower employees to have their interests adequately represented.

References

AlgorithmWatch (2023): Algorithmic Transparency and Accountability in the world of work. A mapping study into the activities of trade unions. https://algorithmwatch.org/en/wp-content/uploads/2023/02/2023_AlgorithmWatch_ITUC_Report.pdf.

Association of Nordic Engineers (2021): Response of the Association of Nordic Engineers to the European Commission's public consultation on the White Paper on Artificial Intelligence – A European approach to excellence and trust. https://nordicengineers.org/wp-content/uploads/2020/10/ane-response-consultation-eu-commission-white-paper-on-ai-2020-final.pdf.

Barocas, S. / Hardt, M. / Narayanan, A. (2019): Fairness and Machine Learning. Limitations and Opportunities. https://fairmlbook.org/pdf/fairmlbook.pdf.

Bolukbasi, T. / Chang, K. W. / Zou, J. Y. / Saligrama, V. / Kalai, A. T. (2016): Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. Advances in neural information processing systems, 29. https://arxiv.org/pdf/1607.06520.pdf.

Bird, S. / Barocas, S. / Crawford, K. / Diaz, F. / Wallach, H. (2016): Exploring or Exploiting? Social and Ethical Implications of Autonomous Experimentation in AI. Workshop on Fairness, Accountability, and Transparency in Machine Learning. https://ssrn.com/abstract=2846909.

Bomassani, R. et al. (2021): On the Opportunities and Risks of Foundation Models. https://arxiv.org/abs/2108.07258.

Campolo, A. / Crawford, K. (2020): Enchanted Determinism: Power without Responsibility in Artificial Intelligence. In: Engaging Science, Technology, and Society, 6, 1–19. DOI: https://doi.org/10.17351/ests2020.277.

Caton, S. / Haas, C. (2020): Fairness in machine learning: A survey. https://arxiv.org/abs/2010.04053.

Christl, W. (2021): Digitale Überwachung und Kontrolle am Arbeitsplatz. https://crackedlabs.org/daten-arbeitsplatz.

Creager, E. / Madras, D. / Jacobsen, J. H. / Weis, M. / Swersky, K. / Pitassi, T. / Zemel, R. (2019): Flexibly fair representation learning by disentanglement. In: International conference on machine learning, 1436–1445. https://arxiv.org/pdf/1906.02589.pdf.

Dobbie, W. / Goldin, J. / Yang, C. (2016): The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges. National Bureau of Economic Research. In: American Economic Review, 108, 201–240. www.aeaweb.org/articles?id=10.1257/aer.20161503.

Ensign, D. / Friedler, S. A. / Neville, S. / Scheidegger, C. / Venkatasubramanian, S. (2018): Runaway feedback loops in predictive policing. In: Conference on Fairness, Accountability and Transparency, 160–171. https://arxiv.org/abs/1706.09847.

European Commission (2020a): White Paper on Artificial Intelligence – a European approach to excellence and trust. https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en.

European Commission (2020b): Artificial intelligence – ethical and legal requirements. Consultation from 20 February 2020 to 14 June 2020. https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements/public-consultation_en.

European Commission (2021a): Annexes to the proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_2&format=PDF.

European Commission (2021b): Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206&from=EN.

Fernando, M.-P. / Cesar, F. / David, N. / Jośe, H.-O. (2021): Missing the missing values: The ugly duckling of fairness in machine learning. In: International Journal of Intelligent Systems, 36, 3217–3258. DOI: 10.1002/int.22415.

Gebru, T. / Morgenstern, J. / Vecchione, B. / Wortman Vaughan, J. / Wallach, H. / Daumé III, H. / Crawford, K. (2021): Datasheets for Datasets. https://arxiv.org/pdf/1803.09010.pdf.

German Trade Union Confederation (2021): The German Trade Union Confederation's Position on the EU Commission's draft of a European AI Regulation. www.dgb.de/downloadcenter/++co++9341cf1a-5107-11ec-9432-001a4a160123.

Gießler, S. (2021): Was ist automatisiertes Personalmanagement? https://algorithmwatch.org/de/wp-content/uploads/2021/05/Was-ist-automatisiertes-Personalmanagement-Giesler-AlgorithmWatch-2021.pdf.

Holubová, B. (2022). Algorithmic management: Awareness, risks and response of the social partners. Final report. Friedrich-Ebert-Stiftung, Competence Centre on the Future of Work. https://library.fes.de/pdf-files/bueros/bruessel/19524.pdf.

Huyen, C. (2022): Designing Machine Learning Systems. Sebastopol, California: O'Reilly Media.

industriALL (2021): Artificial Intelligence: Humans must stay in command. Policy Brief 2019-01. https://news.industriall-europe.eu/documents/upload/2019/2/636849754506900075_Policy%20Brief%20-%20Artificial%20Intelligence.pdf.

Jarrahi, M. H. / Newlands, G. / Lee, M. K. / Wolf, C. T. / Kinder, E. / Sutherland, W. (2021): Algorithmic management in a work context. In: Big Data & Society, 8. DOI: 10.1177/20539517211020332.

Kleinberg, J. / Mullainathan, S. / Raghavan, M. (2016): Inherent trade-offs in the fair determination of risk scores. https://arxiv.org/abs/1609.05807.

Krauth, K. / Wang, Y. / Jordan, M. I. (2022): Breaking Feedback Loops in Recommender Systems with Causal Inference. https://arxiv.org/abs/2207.01616.

Krzywdzinski, M. / Pfeiffer, S. / Evers, M. / Gerber, C. (2022): Measuring Work and Workers. Wearables and Digital Assistance Systems in Manufacturing and Logistics. WZB Discussion Paper, SP III 2022–301. Berlin: Wissenschaftszentrum Berlin für Sozialforschung. https://bibliothek.wzb.eu/pdf/2022/iii22-301.pdf.

LO Sweden (2021): Remiss av Europeiska kommissionens förslag till förordning om harmoniserade regler för artificiell intelligens. www.lo.se/home/lo/res.nsf/vRes/lo_fakta_1366027472949_remiss_eu_k_harmoniserande_regler_artificiell_intelligens_pdf/$File/remiss_EU-K_harmoniserande_regler_artificiell_intelligens.pdf.

Mitchell, M. / Wu, S. / Zaldivar, A. et al. (2019): Model cards for model reporting. In: FAT* '19: Proceedings of the conference on fairness, accountability, and transparency, January 2019, 220–229. https://arxiv.org/abs/1810.03993.

Ponce Del Castillo, A. (2021): The AI Regulation: entering an AI regulatory winter? Why an ad hoc directive on AI in employment is required. ETUI Policy Brief. www.etui.org/sites/default/files/2021-06/The%20AI%20Regulation.%20Entering%20an%20AI%20regulatory%20winter_2021.pdf.

Schelter, S. / Stoyanovich, J. (2020): Taming Technical Bias in Machine Learning Pipelines. In: IEEE Data Engineering Bulletin. https://ssc.io/pdf/taming-technical-bias.pdf.

Syndicom (2020): 9 KI-Leitprinzipien für eine menschenfreundliche Zukunft. https://syndicom.ch/unserethemen/dossiers/kuenstlicheintelligenzki/herausforderungen/.

Trade Union Congress (2022): People-Powered Technology: Collective Agreements and Digital Management Systems. www.tuc.org.uk/sites/default/files/2022-08/People-Powered_Technology_2022_Report_AW.pdf.

UNI Europa (2019): UNI Europa response to the European Commission consultation on AI ethics guidelines. www.uni-europa.org/news/uni-europa-response-to-the-european-commission-consultation-on-ai-ethics-guidelines/.

All websites were last visited on March 13, 2022.

Authors

Dr. Anne Mollen is a senior policy and advocacy manager at AlgorithmWatch as well as the project manager for “SustAIn – Sustainability Index for Artificial Intelligence”. She also works as a media and communication scholar at the University of Münster, where she is researching the interrelation between digital media technologies, society and democracy. The focus of her work lies in automated decision-making systems in the areas of labour, online platforms and the sustainability of AI.

Lukas Hondrich is a research associate at AlgorithmWatch who has been working on Labour Rights and AI Regulation in the EU. He is investigating ways in which workers and trade unions can have more say in data-driven socio-technical systems. Before joining AlgorithmWatch, Lukas developed Machine Learning systems in the e-commerce and tech industries. He holds a Master’s degree in Cognitive-Affective Neuroscience from Technische Universität Dresden and a Bachelor’s degree in Psychology from Johannes Gutenberg University Mainz.

ISSN 2509-2359