Hiring Algorithms: An Ethnography of Fairness in Practice

Introduction

Although AI is claimed to be superior over human experts in terms of objectivity, efficiency, and effectiveness (Hippis 2017; Miller 2015), its growing influence on the outcomes of our daily lives has become increasingly a source of concern. Critics have accused AI of committing ethical violations, such as being racist, sexist, or harming privacy (Crawford and Calo 2016; O’Neil 2016). For example, Amazon recently had to withdraw its automatic hiring algorithm after discovering that it consistently downgraded resumes of female candidates (Dastin 2018). Several agendas have therefore emerged aimed at addressing and solving ethical issues surrounding AI applications, for example, by calling for regulations to enforce transparency, simplifying machine learning models, or devising auditing algorithms that can evaluate discriminatory outcomes (Crawford and Calo 2016; O’Neil 2016; Pasquale 2015; Selbst et al. 2019). However, so far literature offers little insight into how organizations deal with ethical values and AI in practice. Therefore, this research aims to address the following research question: how do organizational groups engage with AI in practice and how are their ethical values being reconfigured?

We address this question by studying an AI application used for recruitment at a large multinational company. The use of AI in the field of Human Resources (HR) is sometimes also referred to as “people analytics”, and relates to the growing trend of data-driven workforce decisions (Gal et al. 2017; Siegel 2013). By allowing managers to make decisions based on objective, measurable “facts”, rather than intuition and subjective experience, AI aims to bring fairness to the traditional ways of managing people in the workplace (Gal et al. 2017; Siegel 2013). AI is assumed to remove subjectivity and bias from the world of work, by presenting decision-makers with a single, evidence-based representation of the truth (Isson and Harriott 2016). In fact, by bringing in a fundamentally different approach for how fair outcomes should be reached, AI introduces new means for organizations to enact ethical values in the workplace. To trace the consequences of AI for ethical values, we take a “practice perspective on values” (Barrett et al. 2016; Gehman et al. 2013), i.e. a study of the specific practices and interactions in the workplace through which ethical values are performed.

We are currently at the stage of iteratively developing theory based on an in-depth study of this phenomenon. In this research-in-progress paper we report on our preliminary findings. We find that AI in recruitment brings to the fore the role of ethical values in organizational decision-making in several ways. First, when AI is not yet used, it may unite organizational groups around its promise of achieving fairness. As groups actually start to implement and work with AI in practice, however, AI exposes radically different means to achieve ethical values which were previously latent. Third, as groups continue to work with AI across a range of situations, AI serves as a trigger for conflict around diverging perceptions about how ethical values should be enacted. We illustrate these findings by our process analysis of the specific practices and interactions of multiple stakeholders in the workplace involved in working with AI.

Method

We conducted an ethnographic in-depth study at the HR department of a large multinational company in Europe, “BeverageCo” (pseudonym), that recently implemented AI to enable fair and bias-free hiring decisions. At the time of data collection, BeverageCo belonged to one of the world’s largest Fast-Moving Consumer Good (FMCG) companies, with annual revenues exceeding $50 billion. The company had almost 200,000 employees in more than 50 countries worldwide. In September 2018, BeverageCo launched an AI application for the recruitment process of all its graduate trainee programs in Europe. Candidates could apply for four different trainee programs in ten different European locations. The trainee programs were highly competitive: yearly more than 10,000 candidates applied for 100 open positions. The AI application was delivered by an external vendor, “NeuroYou” (pseudonym), that promised the organization to remove subjectivity and bias from workforce decisions, by drawing on data science, neuroscience, and machine learning. The AI application replaced the standardized online tests BeverageCo used before (e.g. logical
reasoning test) by neuroscience games, and added the possibility of automated video analysis. Three organizational groups were involved with the AI application at BeverageCo: the HR team, which consisted of multiple recruiters and two HR managers, the AI team, which consisted of several data scientists and a People Analytics (PA) manager, and managers representing the different business units of the company and serving as assessors during selection events.

We have conducted 7 months (726 hours) of non-participant observation of the work around the AI application in graduate recruitment - including 110 meetings and 27 selection events - in the period between October 2018 and April 2019. During our observations, we focused on the design and development of the AI application, including the development and analysis activities of the AI team around the hiring algorithm. Moreover, we focused on the activities of the HR team and managers around the selection of candidates, such as their use of criteria and AI output during selection events. In addition, we also conducted 32 formal interviews and 18 informal interviews with candidates, the AI team, the HR team, and the external vendor. The interviews with the decision-makers around the AI project served to gain understanding of the objectives, context, developments and implications of AI for work. In the interviews with candidates, we asked them to walk us through the specific steps of their selection process, and about their experiences with the AI and human assessment.

We followed a process research approach (Langley 1999) to track the flow of events and understand the unfolding work with AI in practice. In analyzing our data, we started with creating a list of the events at the organization to represent and organize the evidence across phases and identify patterns. We then organized the key events, separating those representing the perspectives of the HR team, AI team, and managers by systematically going through the field evidence for each occupational group. We inductively developed a process model based on the process analysis of the BeverageCo case shown in Table 1. The process model helps in understanding how the HR team, AI team, managers, and candidates engaged in struggles and actions while using the AI application.

Findings

In the following sections we discuss the evolving use of the AI application and the implications for the ethical values of the HR team following the phases summarized in Table 1.

<table>
<thead>
<tr>
<th>Phase 1: Pre-implementation AI Importance of Fairness</th>
<th>Phase 2: Implementation AI Promoting AI to Increase Fairness</th>
<th>Phase 3: Post-implementation AI Questioning Role of AI in Fairness</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Actions by HR team</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Promote fairness</td>
<td>Spearhead AI agenda</td>
<td>Deviate from threshold</td>
</tr>
<tr>
<td>Guard fairness</td>
<td>Pilot AI</td>
<td>Receive candidate feedback</td>
</tr>
<tr>
<td>Teach fairness</td>
<td>Promote AI</td>
<td>Have conflicts with managers</td>
</tr>
<tr>
<td>Have consistent threshold</td>
<td>Find out inconsistency AI</td>
<td></td>
</tr>
<tr>
<td><strong>Actions by managers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adhere to fairness</td>
<td>Participate in AI pilot</td>
<td>Push to hire own people</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oppose diversity promises</td>
</tr>
<tr>
<td><strong>Actions by AI team</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N.A.</td>
<td>Build hiring algorithm</td>
<td>Show inconsistency in AI</td>
</tr>
<tr>
<td></td>
<td>Provide threshold</td>
<td></td>
</tr>
<tr>
<td><strong>Actions by candidates</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N.A.</td>
<td>Participate in AI pilot</td>
<td>Complain unfair AI assessment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Attempt to game the system</td>
</tr>
</tbody>
</table>

Table 1. Process Model of Unfolding AI and Fairness

Phase 1. Pre-Implementation of AI

Before the implementation of the AI application at BeverageCo, members of the HR team widely shared the ethical values of fairness in the recruitment process. They assumed that subjectivity and bias was inherently a problem in human assessment, and considered it as their key task to limit the potential impact of these ethical violations in the workplace. According to the team, fair and credible assessment that allowed for objective measurement of a candidate’s potential, would enable the company to reach the broader aim of increasing diversity in the workplace. As explained by the HR manager: “So why are we looking for objectivity and fairness in the recruitment process? It is part of a wider strategy to make sure that we have
diversity within our company. And not just diversity like gender and nationality, but actually diversity of thought” (HR manager 1). Fair assessment, the HR manager reasoned, would result in a more diverse group of employees in terms of skill set and traits.

HR team members acted as the main promoter and guardian of the ethical values at BeverageCo, and engaged in multiple activities to support these values. During selection events in which managers would assess candidates, the HR team removed resumes because of the potential for bias, gave “unconscious-bias trainings”, and coached managers how to assess fairly. For instance, an HR professional would emphasize before the start of a selection event: “Candidates need to feel they had as much time as the other candidates. That they have been assessed fairly” (Field notes group panel). HR team members also corrected managers when they engaged in biased practices, such as using “lack of culture fit” as an excuse to reject candidates for which they could not explain their reasoning. As illustrated by a comment of an HR manager: “We often use cultural fit as argument [to reject a candidate]. But why? We are going to push you a bit to explain why you think someone does not fit with the company culture” (Field notes in-house day).

Managers were aware of the guarding role of HR and sometimes joked about HR team members acting as “the bias-police”. However, managers seemed to generally agree with the importance of fairness, by learning about sources of bias, correcting each other on subjective judgment, and admitting own potential biases. For example, a manager openly admitted his bias about a stereotypical, extremely polite, old-fashioned British candidate, and asked about the opinion of the other managers: “I was 100% biased. When the guy walked in I shut him out. I am very open here to you guys. Just the whole appearance didn’t work. The way he walked, the way he spoke. […] But I come biased! So, feel free to challenge” (Field notes group panel). In sum, before the introduction of the AI application, the HR team played an important role in promoting, guarding, and teaching the value of fairness to the organization, and in particular to managers that served as human assessors during the recruitment process.

**Phase 2. Implementation of AI**

**Piloting and Developing AI**

The HR director of BeverageCo spearheaded the agenda for the use of AI in recruitment by bringing on board a People Analytics (PA) manager, a newly established role for the purpose of the AI project. Together they started to define a shortlist of AI vendors in the market that would help to achieve HR’s aims of fair and bias-free hiring decisions. Eventually, they decided to enter the pilot stage with vendor “NeuroYou”. NeuroYou offered an AI application involving more than 50 neuroscience gamified assessments and automated video analysis, promising fair and objective measurement of over 120 skills and traits of candidates.

The HR team piloted the AI application in November 2017. The pilot aimed to collect the training data necessary to build the algorithm, identify traits of top-performers, and assess bias in the current hiring process. HR managers collected the data to train the algorithm by asking more than 350 current graduate trainee candidates and 40 top-performing employees to complete the neuroscience gamified assessments and asynchronous video interview via a provided link. Top-performers were identified based on their performance appraisal scores. The neuroscience games aimed to measure cognitive (e.g. “task-switching”), social (e.g. “assertiveness”) and emotional (e.g. “expression recognition”) skills and traits of participants in an easy-to-use interactive environment. In the video interviews candidates were asked to record short answers to questions about themselves, responding to questions such as: “What is your most significant and challenging achievement to date?”. Video interviews helped to extract data on verbal and nonverbal communication, intonation, and facial expressions.

The design and development of the hiring algorithm that would be used for new job applicants was done separately from the team that spearheaded the values: the AI team built the algorithm based on the training data. The training data was taken from the neuroscience games and video interviews completed by the pilot participants. The hiring algorithm extracted the traits of top-performing employees from the training data, and matched these traits against job applicants to identify those applicants with the highest overlap in traits. The algorithm would calculate an algorithmic recommendation per candidate, representing the overlap of the candidate with the top-performer profile at BeverageCo. The HR managers were presented with the final results of a comparative analysis of top-performing employees and candidates by the AI team, but did not receive insight in the specific traits and relative weights assigned to the traits included in the algorithm.
For example, the results showed that candidates were more likely to be hired when they were extravert, while top-performers were less extravert than an average employee. The HR managers concluded: “This suggests bias and shows we are not always assessing candidates on what actually makes them successful”. The HR managers decided based on the pilot results that the AI application was successful in identifying traits and exposing bias in their current hiring process.

In September 2018, the AI application went live for the recruitment of all graduate trainee programs in Europe. Candidates were now completing a four-step selection process, consisting of neuroscience games, a video interview, an in-house day, and a final group panel (see Table 2). The HR managers explicitly decided that human decision-makers would be involved in every step, as they reasoned that human oversight was still required for fair and complete assessment. At the first step, candidates only interacted with AI: they completed 13 neuroscience games measuring different cognitive, social, and emotional skills and traits. HR professionals then manually selected candidates based on the algorithmic recommendation calculated by the hiring algorithm. At the second step, candidates recorded their answers to 10 video questions. At the time of data collection, HR professionals did not employ the algorithmic recommendation for videos because it was not yet reliable. They chose instead to maintain human control over the selection by assessing how candidates presented themselves, using a standardized scoring form. Only at the third step, candidates finally had an opportunity to meet a human. They visited the in-house assessment center of the company, where they went through a series of role-play assessments and a structured interview to demonstrate their social skills, that were assessed by junior managers using a standardized scoring form. Finally, candidates who reached the final step of the selection process were invited to a group panel interview where senior managers assessed their performance and made the final hiring decision.

<table>
<thead>
<tr>
<th>Step 1: Neuroscience Games</th>
<th>Step 2: Video Interview</th>
<th>Step 3: In-house Day</th>
<th>Step 4: Final Group Panel</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI/human assessment</td>
<td>AI</td>
<td>Human/AI</td>
<td>Human</td>
</tr>
<tr>
<td>Online/offline assessment</td>
<td>Online</td>
<td>Offline</td>
<td>Offline</td>
</tr>
<tr>
<td>Final decision-maker</td>
<td>HR team</td>
<td>HR team</td>
<td>Junior managers</td>
</tr>
<tr>
<td>Measured skills and traits</td>
<td>Cognitive, social, and emotional</td>
<td>Social</td>
<td>Social</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 2. The Different Steps of the Selection Process.**

**Using AI to Promote Fairness**

In order to use AI as a means to enact fairness, HR team members started to actively promote the AI application to managers. They argued that the AI application would enable to “objectively measure soft skills and human traits, expose diversity, and compare candidates without bias” (Company documents). These promises were promoted to managers during selection events, as illustrated by the following HR manager: “The beauty of this is that for the first time it can help to objectively assess [soft skills and traits]. It is more or less a proxy for diversity of thought. [AI can help us to answer the question of] how can we start building teams that are more diverse?” (Field notes group panel).

HR team members promoted the AI application to managers by referring to the pilot results. For example, during every group panel, they highlighted the exposed differences in extraversion between top-performers and candidates as an example illustrating bias. These insights were enthusiastically received by managers, who wanted to test their own traits with neuroscience games, and were impressed by the ability of AI to show bias in the hiring process. As commented by a manager: “This brings interesting insights, to challenge our own biases on it [the assessment of candidates]” (Field notes group panel). Moreover, HR team members taught managers how to interpret and use the AI output in their own assessment. For example, during a group panel, an HR professional showed the AI output to the managers in the form of a word cloud representing candidate traits and explained how to interpret and use the output:

“HR professional: So, these are all words that came out of the test, and the larger words are more important.
Manager: So, how do you interpret this word, exactly [points to the word “emotionality”]?
HR professional: Well, you can ask during the panel for example, have you ever experienced a situation where your emotions were in the way?” (Field notes group panel)

Finally, HR team members aimed to use AI as a way to enact fair decisions, by using a fixed threshold for rejecting candidates. This fixed threshold was recommended by the AI team based on the pilot results. During weekly team meetings, HR professionals reviewed the candidates, and selected them based on the algorithmic recommendation. This algorithmic recommendation was represented as a “match percentage score” with the top-performer profile of BeverageCo. For example, if the HR team would observe a score of 95, this implied the candidate’s traits matched the top-performer’s traits for 95 percent. This indicated the candidate was predicted to be a highly successful employee at BeverageCo. By using a fixed threshold for rejecting candidates, HR professionals aimed to be fair in the selection process. As explained by an HR professional: “We need a very structured process, because we are dealing with so many candidates. And everyone is assessed in the same way. We need to be objective”. (Field notes weekly HR team meeting). In sum, the AI application united different organizational groups around its promise of achieving fairness.

**Phase 3. Post-Implementation of AI**

As the HR team started to work with the AI application in practice, however, they became aware that AI exposed different means to achieve ethical values which were previously latent. As different stakeholders started to work with AI across a range of different situations, the HR team was confronted with struggles and conflicts taking place within their own work as well as with other groups about notions of fairness.

**Fairness Struggles of the HR Team**

While HR team members assumed that the AI application would offer them a single, objective, data-driven decision-making point to reject candidates, HR professionals still had to exercise a lot of their personal judgment on when and which candidates to reject. During their work with the AI application on a daily basis, the team experienced that the algorithm did not allow for differentiation between the situated contexts of the different programs, locations, temporary changes in supply and demand, and specific candidate cases. HR professionals therefore increasingly had to make exceptions and changes to the fixed threshold. As explained by an HR professional:

“The threshold depends on the number of candidates that we have. So, in a simple way, how picky we can be. We know for the management program in Germany that we can really hand-pick every single candidate. We have a great brand name there. So, there we decided that the threshold of the match is at least let’s say 85%. For the video interviews, at least 86%. And so on. But for example, for the supply program in the UK we know that we cannot be so picky.” (Field notes weekly HR team meeting)

HR professionals deviated from the threshold based on the supply and demand of candidates, specific candidate requirements (e.g. language, educational background), or personal arguments for selecting candidates. For example, when a candidate had a score slightly below the threshold, HR professionals would often say: “Can we give him a chance?” or “If it’s one percent difference, you can just move her forward”. These deviating practices raised concerns of the HR manager, who feared the changes undermined fairness in the selection process. The HR manager started scheduling meetings with the AI team to decide on the thresholds she should set. She explained this priority to the AI team:

“So, what we started implementing as an approach this year was differentiating thresholds, and I really want to look into the details of thresholds to have the right ones for the entire [recruitment] cycle, with no changes. So, that’s the priority for me, to have a super fair process during the [recruitment] cycle.” (Field notes HR and AI team meeting)

**Fairness Struggles of Candidates**

Second, while the HR team members hoped that the AI assessment would result in a fair selection process for all candidates, they found out that candidates’ experiences with fairness could differ. HR managers were confronted with several candidates who responded negatively about a fair selection process during recruitment events, selection events, or via email. For example, candidates expressed they did not recognize themselves in the AI output they could access after completing the AI assessment. During one specific
instance, the HR manager received an email of an upset candidate, stressing that his results were inaccurate and demanding that his data was deleted. The HR manager complied with this request. Other candidates complained they felt unfairly treated because they lacked the opportunity to perform, as they had the impression they were just “playing a game”. This resulted in the HR manager worrying about how candidates experienced fairness of the selection process.

In contrast, HR professionals were also confronted with candidates who aimed to gain an unfair advantage over other candidates in the selection process by “gaming the system”. For example, HR professionals found out that several candidates aimed to bypass the system by creating a new account with a different email address, in the hope to improve their AI scores. An HR professional addressed one instance during a weekly team meeting: “So, he was applying for the leadership program, I think, and he had a really low match score. So, we rejected him. And he was so cheeky that he started a second account with another email!” (Field notes weekly HR team meeting).

Although the candidate showed high scores the second time he took the test, the HR team decided to reject the candidate as he had an unfair advantage over other candidates. The possibility to “game the system” was confirmed by candidates during the interviews, in which they explained that it was possible to cheat on several neuroscience games as well. For example, a candidate said about a specific game in which one had to memorize multiple changing figures: “You could actually cheat on those games. If you would do the game with three people, hold your phone in your hand, and both make a picture [of the figure you have to memorize], I am sure you would pass the game” (Candidate 1). Thus, HR team members were confronted with candidates who held different fairness perceptions and strategies that counteracted fairness.

**Fairness Struggles of Managers**

Third, AI conflicted with alternative perceptions of what was considered fair by managers. In several instances, HR professionals were confronted with frustrated managers who could not hire their preferred candidate because the candidate was rejected by the AI application. For example, a sales manager was prevented from hiring his current intern, because the intern “failed” the AI assessments. The intern commented on the incident: “I had a score of 30%, so basically I cannot concentrate. Which is kind of true. But I am lucky that I can still do my internship, because for this internship the test was not required”. The sales manager was furious, and felt it was wrong to prefer AI assessment over four months of human assessment during the course of the internship: “What’s a better selection tool – four months of assessment versus some stupid games! I am quite sure I would also not get through the tests if I would have to do them now. He scores low on concentration, but I fail in other things” (Manager 3).

This resulted in several conflicts between HR team members and managers, in which managers preferred their personal judgment over AI assessment, which was bias according to the HR manager. The HR manager aimed to respond to the critiques of AI assessment by defending its credibility based on the involved people in the development of the algorithm. However, this was often unsatisfactory to managers, as illustrated by this example:

> “HR manager: We created the algorithm based on last year’s decisions. This is the way we built it. Then you would question the presence of all of the graduates and all the other people that collected the data for the algorithm.
> Sales manager: Yeah, but still our human assessment should be leading. If the person is super great and nails that but doesn’t pass the AI assessment, how do you explain that?”
> (Field notes office area)

In addition, HR team members were confronted with managers who were puzzled about how exactly the AI application would enable achieving the diversity aims. For example, when an HR professional explained the AI assessment to managers during a group panel, one of the senior managers objected: “We will have less diversity because we will hire more of the same profile, right?”. Another manager expressed the fear of “cloning people”, by selecting candidates on the same set of traits. Several managers started to engage in actions to counteract the homogenous focus of the AI assessment, by focusing on other traits in their human assessment. This was illustrated by the following discussion between two managers:

> “Manager 1: We have to watch out that we don’t get only leaders.
> Manager 2: Yes, but with AI we are only selecting similar profiles. We basically don’t have any diversity.
Manager 1: I always assess if people take the lead, but also if they let others speak.”
(Field notes in-house day)

In sum, HR team members started to experience several critiques of managers who held alternative perceptions of fairness, and started to criticize or oppose the AI assessment.

**Fairness Struggles of the AI Team**

Finally, while HR team members assumed that the use of AI resulted in objective recruitment outcomes, the team was confronted by the AI team with inconsistent use and outcomes of the AI application. The HR team and AI team had several team meetings in which the AI team would present the latest analysis results to the HR team. During a team meeting in February 2019, the PA manager explained to the HR manager an inconsistency in their use of the hiring algorithm: not all candidates were assessed and selected based on the algorithmic recommendation, with one group automatically passing to the next step of the selection process, thereby escaping the algorithmic filter. This inconsistency was problematic for the PA manager as it implied that the obtained data was in fact contaminated and rigorous comparative analysis on the candidates before and after the use of the algorithm was compromised. She emphasized to the HR manager: “This is a challenge for me, because the assessment was not fair there” (Field notes meeting HR team and AI team). Thus, the HR team became aware that not all candidates were subject to the same hiring algorithm, pointing to violations of fairness in selection. Moreover, the inconsistency in the use of the hiring algorithm showed to HR team members that surprisingly some candidates did in fact receive hiring offers from managers, despite the algorithm predicting the opposite. As explained by an HR professional:

“So, Alex, this is the only one guy who got the offer for the sales program in Spain. He declined it, but he got it. With a video score of 69, and a game score of 83. But actually, he was not assessed by the algorithm. [...] And actually, right now, he would be rejected. And he is the only one guy that got the offer. So, candidate’s success is unpredictable actually, I would say.”
(Field notes weekly HR team meeting)

This observed gap between the algorithmic recommendation and hiring outcome resulted HR team members losing trust in the AI application. HR professionals therefore increasingly started deviating even more from the thresholds when working with the AI application, arguing: “Because we don’t know what [score] works and doesn’t work basically” (Field notes weekly HR team meeting). During a team meeting in March 2019, the doubts of the HR team culminated: while the PA manager presented the latest analytic results, the HR manager criticized and questioned the validity of the results. The results showed that top-scoring candidates (i.e. candidates with match scores of 92%) were rejected at the human steps of the selection process and did not receive any hiring offers. This resulted in a fierce discussion between the HR manager and PA manager:

“HR manager: Maybe it is also a question about whether this [AI assessment] actually works?
AI manager: You mean, do the in-house days work? [...] HR manager: But in-house days or the AI assessments, right? It can be one or the other failing. They are not matching.” (Field notes meeting HR and AI team)

While the HR manager questioned the AI assessment, the PA manager responded by blaming the human assessors that were involved in the selection process; human assessors were rejecting predicted successful candidates. The HR team and AI team therefore started to take opposing sides in considering AI as means to enact the fairness values in the recruitment process.

**Conclusion**

Implementing and using AI in recruitment brings to the fore issues around ethical values in organizational decision-making. Our process analysis of the specific practices and interactions of multiple stakeholders in the workplace shows that ethics with AI can take a very different shape from what it promised, when put into practice. In particular, before the use of the AI application, different organizational groups united around the fairness promises of AI. However, as groups started to unpack and work with AI in practice, the various stakeholders became aware of their diverging perceptions of fairness. Our preliminary findings thus illustrate that while organizational groups can discursively agree upon AI as ways to enact ethical values, these values become reconsidered and negotiated once people start to interact with and put AI into practice.
While the literature on ethics and algorithms has so far mostly focused on discussing relevant guidelines on how ethics should be ensured in design (Crawford and Calo 2016; O'Neil 2016; Pasquale 2015; Selbst et al. 2019), our case adds to this discussion by providing an in-depth empirical account of how such ethics is performed in practice. Our ethnographic story allows recognizing that even when ethics is a guiding aim behind AI implementation, how these values are enacted are difficult to conceive upfront, especially when multiple stakeholders are involved.

Previous research has emphasized how work gets reconfigured as a result of algorithmic technologies increasing people’s dependencies on machines (Newell and Marabelli 2015), transforming the standards for valuation of work (Orlikowski and Scott 2014), or enabling power shifts in work relations (Barbour et al. 2018). Our findings extend this line of research by demonstrating the importance of AI in triggering negotiation of ethical values. Our study also builds on research that takes a “practice perspective on values” (Gehman et al. 2013); (Barrett et al. 2016), by demonstrating that ethical values cannot be treated as given, but are socially and materially constructed when developing and using algorithmic technologies in practice. Beyond implications for future research, our preliminary findings also provide an in-depth case of AI in practice that serves as a caution to those who have too high expectations of AI as ways to promote fairness in organizations. In the full paper, our theoretical framework that is currently under development will be included.

References


