Implications of AI (Un-)Fairness in Higher Education Admissions

The Effects of Perceived AI (Un-)Fairness on Exit, Voice and Organizational Reputation

Frank Marcinkowski
Department of Social Sciences
University of Düsseldorf, Germany
Frank.Marcinkowski@hhu.de

Kimon Kieslich
Department of Social Sciences
University of Düsseldorf, Germany
Kimon.Kieslich@hhu.de

Christopher Starke
Department of Social Sciences
University of Düsseldorf, Germany
Christopher.Starke@hhu.de

Marco Lünich
Department of Social Sciences
University of Düsseldorf, Germany
Marco.Luenich@hhu.de

ABSTRACT

Algorithmic decision-making (ADM) is becoming increasingly important in all areas of social life. In higher education, machine-learning systems have manifold uses because they can efficiently process large amounts of student data and use these data to arrive at effective decisions. Despite the potential upsides of ADM systems, fairness concerns are gaining momentum in academic and public discourses. The criticism largely focuses on the disparate effects of ADM. That is, algorithms may not serve as objective and fair decision-makers but, rather, reproduce biases existing within the respective training data. This study adopted a different approach by focusing on individual perceptions of fairness. Specifically, we looked at two different dimensions of perceived fairness: (i) procedural fairness and (ii) distributive fairness. Using cross-sectional survey data (n = 304) from a large German university, we tested whether students' assessments of fairness differ with respect to algorithmic vs. human decision-making (HDM) within the higher education context. Furthermore, we investigated whether fairness perceptions have subsequent effects on three different outcome variables, which are hugely important for universities: (1) exit, (2) voice, and (3) organizational reputation. The results of our survey suggest that participants evaluated ADM higher than HDM in terms of both procedural and distributive fairness. Concerning the subsequent effects of fairness perceptions, we find that (1) distributive fairness as well as procedural fairness perceptions have a negative impact on the intention to protest against an ADM system, whereas (2) only procedural fairness perceptions negatively affect the likelihood of exiting. Finally, (3) distributive fairness, but not procedural fairness perceptions have a positive effect on organizational reputation. For universities aiming to implement ADM systems, it is crucial, therefore, to take possible fairness issues and their further implications into account.

FAT* '20, January 27–30, 2020, Barcelona, Spain

ACM ISBN 978-1-4503-6936-7/20/02…$15.00
https://doi.org/10.1145/3351095.3372867

CCS CONCEPTS

• **Human-centered computing** → *Empirical studies in HCI*; • **Applied Computing** → *Sociology*; • **Applied Computing** → *Education*

KEYWORDS

Distributive Fairness, Procedural Fairness, Artificial Intelligence, Algorithmic Decision Making, Higher Education Systems, Reputation, Voice, Exit

ACM Reference format:

Frank Marcinkowski, Kimon Kieslich, Christopher Starke, and Marco Lünich. 2020. Implications of AI (Un-)Fairness in Higher Education Admissions: The Effects of Perceived AI (Un-)Fairness on Exit, Voice and Organizational Reputation. In *FAT* '20: Conference on Fairness, Accountability, and Transparency (FAT* 20), January 27–30, 2020, ACM, Barcelona, Spain,* 9 pages. https://doi.org/10.1145/3351095.3372867

1 ARTIFICIAL INTELLIGENCE IN HIGHER EDUCATION SYSTEMS

Algorithmic decision-making (ADM) is being applied increasingly in virtually all sectors of social life, such as medicine [23,36], public administration [38], and finance [21]. In addition, institutions of higher education, such as universities, have begun implementing artificial intelligence (AI) applications to predict student performance, analyze academic teaching, communicate with students through the use of bots, perform dropout detection and structure their financial organization [4,11,14,31,40] (for an overview of the use of AI applications in German universities, see [24]). Even the possibility of university admission [1,9] was recently part of the public discourse:

> *It's easy to imagine, for example, how a college-admissions committee might turn the laborious and uncertain sifting of applicants over to a machine-learning model; such a model might purport to optimize an entering cohort not just for academic success but also for harmonious relationships and generous alumni donations* [41].

Such a system could be quite valuable as universities spend considerable financial and human resources on the admission process. In Germany, many study programs at public universities are freely accessible and free of charge. However, it is not always possible to enroll directly in a program at one's university of choice. If there are fewer courses of study available than are requested by interested parties, admission is restricted. At the time of this study, about 40% of the programs nationwide were admission-restricted [18], with the rest being admission-free. Courses in medicine, pharmacy, veterinary medicine, dentistry, and communication science are so popular that they are restricted at all universities nationwide. In addition, a large number of degree programs at individual universities have restricted admission. At the university where we conducted this empirical study, these also include programs, for example, in political science, economics, and sociology. In principle, restrictions apply to study programs that grant access to professions with a particularly high income or that open up access to popular occupational fields (e.g., media business).

The principle of "best selection" generally guides the selection process at higher education institutions; that is, the primary concern is about selecting those applicants who possess as many of the necessary skills and abilities as possible. In a broader sense, the aim is to keep the dropout rate as low as possible because high dropout rates not only damage the reputation of study courses and degree programs but also harm institutes financially, as their public funding is dependent, among other things, on the number of graduates they produce. Applicants with the best average grades on their school reports fill most of the study spaces. In addition, universities can define further selection criteria for local admission procedures, such as selection interviews, internships, weighting of certain grades in the diploma, or a completed study ability test. These criteria differ from university to university and from subject to subject. Since selection interviews, evaluations of motivation letters, or aptitude tests are very resource-intensive, it appears logical to consider the use of automated procedures with AI in the admission process. The use of ADM has the potential to make this process more cost-effective and faster, which could benefit both the university and the applicants [40].

Yet a controversial discourse also exists regarding the fairness of student admission systems in general. On the one hand, it includes the question of which data or information to account for and whether this information is valid. On the other hand, it addresses individual fairness perceptions of decision-making practices and decision outcomes. Thus, we conceptually distinguish *two faces of fairness*. While *factual fairness* refers to objectively measurable features, *perceived fairness* is a construct that relates to individual *perceptions*. Factual and perceived fairness are presumably correlated, but conceptually distinct. Despite the scant empirical research within the social sciences [5], fairness must be treated as a pivotal value with respect to ADM. In a survey, Araujo et al. [3] found that fairness was the second most important value for ADM among Dutch citizens. Given the importance of fairness, it also seems to be of interest to evaluate the outcomes of perceived fairness or unfairness of ADM. Considering that AI-driven systems take over tasks that humans previously executed, it is crucial to investigate differences between ADM and HDM with regard to perceived fairness.

Thus, this study addressed two research questions. First, we investigated whether ADM is perceived as fairer or more unfair than HDM in terms of procedural and distributive fairness. Second, we tested the effects of AI fairness perceptions on attitudes and behaviors relevant to the university organization—namely, exit, voice, and organizational reputation. In particular, we examined the effects of perceived procedural and distributional AI fairness on the intentions to protest ADM and refrain from applying to a university that uses ADM in the application process. Furthermore, we examined the link between AI fairness and the overall reputation of a university that uses such a system.

2 THE CONCEPT OF AI FAIRNESS

In the current literature on ADM, the term "unfairness" usually denotes ADM systems that systematically discriminate against individuals or groups while favoring others. This would be the case, for example, if ADM systems were to reject university applicants because they have a Muslim sounding name. Since discrimination becomes evident in the result of an automated classification, this is referred to as a violation of distributive fairness [10]. To be able to classify a certain ADM outcome as unfair, we require the violation of a valid norm by the distribution result. For instance, the constitutions of all democratic states prohibit discrimination against people based on their race, gender, religious beliefs, etc. An ADM system that does just that is obviously unconstitutional and factually unfair. The task of computer science, then, is to adapt the algorithm such that it can avoid this norm violation without impairing the performance of the ADM. Insofar as this is demonstrably successful, *proof of fairness* is achieved. Similarly, it would be possible to confirm an objective breach of procedural fairness if the ADM were to access legally protected data that the service provider was not allowed to use. In both cases, unfairness would be measured against objective criteria that violate a given standard. Studies on such phenomena point to the need for political and legal regulation of the use of ADM in a given area of application [7,12,20].

Our interest was not in taking the perspective of decision-makers, lawyers, or politicians; rather, we were interested in the point of view of those affected by the use of ADM within an institution. In other words, we concentrate on perceived fairness rather than factual fairness. We based this research on the assumption that factual and perceived fairness of algorithmic systems are equally important: It is not enough for technical systems to have been tested and determined to be fair and working efficiently if people nevertheless encounter them with fear and mistrust and subsequently become demotivated and dissatisfied, and lose their attachment to the respective institution. It is also not enough for people to be satisfied with technical systems even though they violate objective requirements for fair decision-making, legal norms, or ethical rules. While technical definitions and formal proof of fairness are widely acknowledged within the machine learning community, it is time to focus on individual perceptions [22]. Individual reactions include emotional (e.g., fear), cognitive (e.g., trust), and behavioral (e.g., protest) responses of different kinds. We assume that the perceived fairness of ADM strongly influences human responses to ADM. Perceived fairness is not (only) about objective norm violation but also about perceived violations of the individual sense of justice by ADM systems. This user-centered approach [25,29,33] views the subjective perceptions of the treatment that an individual receives from other individuals or institutions as the core of the concept. Therefore, perceived fairness is an individual cognitive reaction directed towards manifest actions of third parties. Shin and Park [33:283] conclude that "There might not be an objective standard of transparency, fairness, and accountability. The subject of FAT [fairness, accountability, and transparency] lies in the eye of the beholder." Of course, the perceived fairness of an ADM is not independent of factual fairness, but it is also not identical to it. If an algorithm classifies without discrimination in the aggregate, it can still be perceived as unfair at the individual level. Since the individual user normally does not know which distribution result an ADM produces in the concerned population (e.g., who gets a loan from the bank and who does not; who is classified as a potential dropout and who is expected to graduate), there is no other option but to base the assessment on one's own results. In the worst case, those affected will always assume that they have received an unfair share of the burden to be distributed and the benefits to be gained when the automated classification result has fallen short of their expectations. Subjective perceptions also play a role concerning procedural fairness, which is usually determined by computer scientists on the basis of the quality of data used [17]. For example, someone affected by an automated recruitment process will perceive the result as fair if he/she has reason to believe that the information used in the decision-making process was correct, reliable, valid, and complete. However, a rational judgment is hardly possible, which is why the perceived procedural fairness will be based on more or less well-founded assessments, prejudices, and assumptions.

A concept of perceived fairness understood in this way, obviously, cannot be determined definitively; rather, the perceived fairness of ADM technologies is always specific. The respective assessment dimensions used and individual assessments of these categories can vary considerably from person to person depending on the area of application and use case [34]. Social science research is aimed at determining on which factors individual perceptions of fairness depend and what consequences this might have [2,3,8,26,30]. On the side of independent variables, personality factors as well as characteristics of the technology itself, but also characteristics of the situation and use case, are taken into consideration. Typical dependent variables are acceptance of ADM decisions, trust in their correctness, but also behavioral responses, such as voice or exit.

3 DISTRIBUTIVE AND PROCEDURAL FAIRNESS

3.1 Concepts

The organizational justice literature suggests using a multidimensional construct of fairness [10,17]. In this paper, we focus on two of those dimensions—procedural and distributive fairness. In that sense, it is possible to judge fairness by the process—that is, in terms of machine learning the input and throughput phases—and the result of an AI-driven process (output). Arguably, fairness judgments can differ between those dimensions [17].

Distributive fairness refers to the fair distribution of resources [39]. Verma and Rubin [37] provided an overview of the mathematical formulae ensuring a fair outcome. In the case of ADM, algorithmic distributive unfairness may be the outcome of biased training data that (unwillingly) reproduce real-world discrimination. In Germany, for example, children of workers could be denied access to higher education simply because they are currently underrepresented at universities compared to middle-class children [6,20,35]. Discrimination may then lead to perceptions of distributive unfairness. However, it is possible to achieve distributive fairness if one perceives that the outcome of an AI process is just and fair. Distributive fairness perceptions are dependent on the respective individually accepted distribution norms at hand [34], which can vary considerably between different contexts. For instance, while we might expect basic human rights to be distributed according to the principle of equality, salaries or social benefits are typically distributed on the basis of equity or need [13]. Some studies have looked into the determinants of distributive fairness perceptions and have found that process-related variables impact how people evaluate the outcome of ADM [8,16].

Procedural fairness relates to the process, or in technical terms, the data and mechanism used to achieve the outcome. According to the literature on HDM, procedural fairness can be evaluated by considering six factors: (1) consistency, (2) neutrality, (3) precision, (4) revocability, (5) ethics, and (6) representativeness [27]. If we assume that machines are judged against the same standards, ADM will be perceived as fair if people get the impression that the system meets these same criteria.

In AI research, questions of procedural justice are dealt with far less frequently than problems of distributive justice. However, Grgić-Hlača et al. [17] conducted a study on feature selection of ADM tools. They developed a survey asking participants to rate the possible features of an algorithm in three categories: (1) if it is fair to use a feature, (2) if it is fair to use a piece of information if it increases the accuracy of the outcome, and (3) if it is fair to use a piece of information if one group of people is more likely to be falsely predicted [17:54]. The results of the study suggested a huge variety in fairness judgments of features. Features that have a clear thematic linkage to the aim of the algorithm tend to be rated as fair. On the contrary, demographic features, like age, gender, or race, tend to be regarded as the most unfair features—even if those features may enhance the precision of an algorithm. Thus, judgments of the fairness of an algorithm appear to depend on the imagination and selection of the feature. Furthermore, according to Grgić-Hlača et al. [16], features are not only rated as fair on the basis of discriminatory impact. Fairness judgments are built on criteria like the reliability and relevance of the feature or if it causes a specific outcome.

3.2 ADM vs. HDM

Oftentimes, AI applications take over tasks that humans previously executed. Thus, there is a growing body of literature comparing the perceptions of human and algorithmic decision-makers. In a series of experiments, Logg, Minson, and Moore [28] identified that laypersons in particular show an "algorithmic appreciation"; in other words, they ascribe many positive characteristics to algorithms and trust ADM more than HDM.

Moreover, among other variables, such as risk, trust, and emotions [2,26], fairness is also addressed as a relevant construct with respect to comparing ADM and HDM. The research suggests that differences in fairness perceptions depend on the respective task that is carried out [2,26]. One study found no difference between evaluations of the fairness of humans and machines in tasks requiring mechanical skills. However, the authors did find that tasks requiring human skills were, in fact, rated as fairer when humans executed them [26]. By contrast, another research team found that in high impact situations in the context of justice and health decisions, AI was perceived as fairer than humans experts [2].

As we may regard admission to a study program as a high impact situation for a student and the application process as a merely technical task, we propose following hypotheses:

H1a: ADM applications are attributed higher distributive fairness than HDM committees are.

H1b: ADM applications are attributed higher procedural fairness than HDM committees are.

4 IMPLICATIONS OF PERCEIVED (UN−)FAIRNESS IN THE HIGHER EDUCATION CONTEXT

As stated earlier, we propose that fairness perceptions trigger specific reactions. In our study, we focused on two behavioral intention variables—exit and voice—as well as one cognitive variable—reputation.

As public institutions, universities in Germany receive public funds according to the number of enrolled students. Consequently, they compete with each other over student applications. In this regard, a positive reputation among current and future students becomes a pivotal resource for public universities. Should the impression arise at the very beginning of a student career—namely, during the application process—that the university treats applicants unfairly, presumably, this will not only influence the students' behavior but might also damage the university's reputation overall. For instance, the recent college admissions bribery scandal in the United States makes it plausible to assume that a defective and unfair admission process may have substantial detrimental effects on the university's reputation. To measure students' reactions to a university deploying ADM systems, we used Hirschman's [19] seminal distinction between exit and voice. While neoclassical theory assumes that customers who are dissatisfied with a provider's services simply switch to a competitor—that is, exit—Hirschman stressed that under the conditions of limited competition in monopolized markets, other modes of response are more likely and probably also more effective: Clients articulate their protest with the aim of changing grievances. In addition to those behavioral variables, we assessed students' cognitive responses with respect to the perceived reputation of a university using ADM. This leads to the following hypotheses:

H2: Perceived distributive and procedural AI fairness negatively affect the intention to move to another university (exit).

H3: Perceived distributive and procedural AI fairness negatively affect the intention to voice one's protest.

H4: Perceived distributive and procedural AI fairness positively affect the university's reputation.

5 METHOD

5.1 Procedure

To test our hypotheses, we conducted a laboratory survey among 304 students of a large university in Germany between June 17 and 28, 2019. To recruit participants, we attached posters to notice boards in several public university buildings as well as personal recruitment on the university campus. Personal recruitment took place at all highly trafficked spots on campus, such as the cafeteria and the library, to obtain a reliable sample of the student population of the university. Therefore, it is worth noting that our study was susceptible to the limitation of self-selected samples. Although we were prudent about endeavoring to reach out to all students, it is plausible to assume that some students were a priori excluded from the sample (e.g., students who were not on campus when the study was conducted).

The study was advertised under the neutral title "University of the future" to minimize selection bias. Students were asked to sign up for a date online or to drop by the laboratory spontaneously.

Participants first had to sign a privacy statement before completing the questionnaire. The average time to complete the questionnaire was approximately 15 minutes (M = 14.48, SD = 3.83). To begin with, participants answered some questions regarding their attitudes toward, and knowledge of, digitalization. Next, they read a short explanation of how human committees currently decide university admissions to programs with limited spaces. The questionnaire pointed out that AI technology could take over the task by analyzing all available applicants' data and, based on the results, recommend approvals or declines of the applicants. Participants then had to rate the fairness of the HDM and ADM. Following this, they were asked to answer some questions about the reputation of a university that might use such an automated system as well as their likelihood of applying to another university that does not use ADM or protesting the use of such systems at their university of choice. Finally, participants provided demographic information as well as some student information, like the study program in which they are enrolled or number of semesters that they have already completed. Finally, participants were thanked and incentivized with €5.

5.2 Sample

In total, 305 participants completed the survey. However, one case was excluded from the dataset because the participant answered the questionnaire in an unrealistic amount of time (t = 5 minutes). This minimum time-mark was assessed prior to data collection in a pre-test with independent raters. Thus, our final sample consisted of 304 participants. Considering gender distribution, 177 (58.2%) of the participants were women and 124 (40.8%) were men, whereas three participants (1.0%) did not indicate their gender. The vast majority of participants (88.8%) were born in Germany. On average, they were 23 years old ($M = 23.29$, $SD = 4.00$) and had studied for four or five semesters thus far ($M = 4.64$, $SD = 3.15$). Most participants were enrolled in an undergraduate program (67.1%), followed by a graduate program (31.9%), and a PhD program (2.0%). Furthermore, 145 (47.7%) participants were enrolled in the faculty of philosophy, 69 (22.7%) in the faculty of mathematics and natural sciences, 43 (14.1%) in the faculty of economics, 33 (10.9%) in the faculty of medical and health sciences, and 12 (3.9%) in the faculty of law.

The sample depicts the student body of the university quite well in terms of gender distribution (women: 57.6% [university] vs. 58.4% [sample]; men: 42.4% [university] vs. 40.7% [sample]). However, we oversampled students of humanities and social sciences and undersampled students of natural sciences. This might be because the study was conducted in a building belonging to the faculty of philosophy; it is reasonable to assume that many participants who spontaneously took part in the survey were students of the corresponding faculty.

Furthermore, we compared our descriptive data with the official student body statistics in Germany provided by the Federal Statistical Office [15]. In terms of gender distribution, women are overrepresented in our sample by 11%. Additionally, we oversampled students from the philosophy/law/economics departments by approximately 14%, from the mathematics/natural sciences departments by 12%, and from the medicine department by 5%. The oversampling of all disciplines occurred because the university where we conducted the study has no engineering department. As 27% of German students are enrolled in an engineering degree program, our sample inevitably overrepresents all disciplines compared to the official student statistics.

5.3 Measurement

Fairness Perceptions. We measured fairness perceptions for both the HDM committee and the ADM application in terms of distributive fairness and procedural fairness. The former dimension assessed fairness regarding the potential outcome of the admission process. The latter dimension measured the perceived fairness of the process itself. We gauged both dimensions via a single item on a five-point Likert scale (“1 = do not agree at all” to “5 = totally agree”). For distributive fairness perception, participants rated the statement “The selection of applicants is fair” ($M_{HDM} = 2.97$, $SD_{HDM} = .84$; $M_{ADM} = 3.42$, $SD_{ADM} = .99$); for procedural fairness, they rated the statement “The procedure is unbiased” ($M_{HDM} = 2.36$, $SD_{HDM} = .96$; $M_{ADM} = 4.16$, $SD_{ADM} = 1.09$).

Exit. To measure the intention to withdraw from a university that uses an AI-based admission system, we decided to insinuate a negatively associated scenario. Thus, we told participants to imagine that an ADM system declined their application. We adopted this approach from Binns et al. [8].

Respondents then rated the following four items on a five-point Likert scale (“1 = do not agree at all” to “5 = totally agree”): “I'm avoiding the university that turned me down”; “I prefer other universities that do not use an ADM admission system”; “I want nothing more to do with the university that rejected me”; and “I would not apply to a university that uses an ADM admission system.” Subsequently, we computed a mean index for exit behavior, Cronbach's $\alpha = .751$ ($M = 2.75$ $SD = 1.03$).

Voice. We gauged raising one's voice against a university that uses an ADM admission system under the same conditions as exit behavior. Likewise, we measured this via four items on a five-point Likert scale (“1 = do not agree at all” to “5 = totally agree”). The item wordings were as follows: “I would actively oppose the admission system”; “I would participate in a demonstration against the admission system”; “I would sign a petition against the admission system”; and “I would support a protest against the admission system.” We calculated a mean index for protest behavior, Cronbach's $\alpha = .914$ ($M = 2.62$ $SD = 1.17$).

Reputation of the university. We employed the RepTrak^{TM} Pulse to assess the reputation of a university using an AI-based admission system [32]. It consists of the four items “It is a university I have a good feeling about",”; “It is a university that I trust”; “It is a university that I admire”; and “The university has a good overall reputation.” Participants rated these statements on a five-point Likert scale (“1 = do not agree at all” to “5 = totally agree”). We used these four items to compute a reliable mean index, Cronbach's $\alpha = .872$ ($M = 3.09$, $SD = .83$).

6 RESULTS

To answer H1a, we performed a dependent t-test comparing perceptions of distributive fairness between ADM and HDM. Our results show that participants perceived the ADM system ($M = 3.42$, $SE = .057$) as significantly fairer than the HDM system ($M = 2.97$, $SE = .048$), $t(303) = 7.266$, $p < .05$, $r = .38$. That is, regarding the output of the admission process, they considered ADM to be less biased, in one direction or the other, compared to human committee members making the decision. Thus, our data support H1a.

As for H1b, we carried out a dependent t-test to determine differences between ADM and HDM with respect to procedural justice. On average, participants viewed the AI-driven process as fairer ($M = 4.16$, $SE = 0.062$) than the HDM process ($M = 2.36$, $SE = 0.055$), $t(303) = 20.890$, $p < .05$, $r = .77$. This result supports H1b.

Addressing hypotheses H2–H4, we ran three multiple ordinary least squares (OLS) regression models. We entered the predictors in a two-step process. First, we tested the effects of the demographic and study-specific variables and the distributive and procedural fairness perceptions of HDM[1]. In the second step, we entered the perceptions of distributive and procedural fairness concerning ADM[2]. Thus, we account for the proportion of explained variance in the models. Tables 1–3 depict the results.

The first regression model (H2) explains 4.4% of the variance in the dependent variable *exit*, meaning that the explanatory power of the model is rather low. Table 1 depicts the predictors' influence on the dependent variable. We see in block 1 that faculty membership exhibits a significant negative effect. The perception of procedural fairness of HDM also exerts a significant negative effect.

Table 1: OLS regression with dV exit

dV: Exit				
		b	SE	β
Step 1	Intercept	2.428	0.476	
	Gender (1=male)	−0.011	0.113	−.006
	Age	0.032	0.017	.125#
	Faculty	−0.371	0.140	−.152**
	Graduation status	−0.022	0.125	−.010
	Semester count	−0.032	0.022	−.099
	HDM DF	0.072	0.072	.059
	HDM PF	−0.159	0.063	−.148*
Step 2	Intercept	3.139	0.551	
	Gender (1=male)	−0.022	0.113	−.011
	Age	0.028	0.017	.110
	Faculty	−0.328	0.141	−.134*
	Graduation status	−0.050	0.125	−.023
	Semester	−0.030	0.022	−.093
	HDM DF	0.129	0.076	.106#
	HDM PF	−0.181	0.063	−.169**
	ADM DF	−0.054	0.065	−.052
	ADM PF	−0.129	0.056	−.137*
Note: *adj.* $R^2 = .044$, $\Delta R^2 = .017$. $^{\#}p < .10$, $^{*}p < .05$, $^{**}p < .01$				

Turning to the final model, we find that both effects are equally strong ($\beta_{faculty} = -.13$; $\beta_{HDM\ PF} = -.17$). Thus, studying a mathematics or natural science subject lowers the intention to migrate to a different university. Furthermore, perceiving HDM as biased/unfair strengthens the intention to exit. Contrary to our assumptions, perceiving the admission process by HDM as unfair is not necessarily seen as a threat but more as an opportunity. One plausible explanation may be that some students who rate the process as unfair might turn to such universities because they expect their chances of admission to be greater. Turning to fairness perceptions with respect to ADM, the results suggest that ADM procedural fairness also has a negative effect on withdrawal ($\beta = -.14$). Thus, when one perceives the ADM admission process as unfair, one might turn to other universities that do not use an AI-driven system.

Interestingly, distributive ADM fairness perceptions show no significant effect on intention to exit. Thus, it is not the experience of an unfair result that deters applicants. Instead, the mere fact that an ADM system makes the decision is sufficient to persuade potential applicants to move to another university. Therefore, our data partially support H2 owing to the significant effect of the procedural ADM fairness dimension, but not the distributive ADM fairness dimension.

Overall, the second model explains 16.2% of the variance of the dependent variable, whereas the lion's share is explained by the ADM-specific fairness variables ($\Delta R^2 = .11$). Table 2 depicts the regression coefficients for the predictors of the dependent variable *voice*. Age and faculty membership both have a significant effect on *voice*. Thus, older students are somewhat more sympathetic to protesting in the case of an ADM system declining their university application. We further find that studying in the mathematics or natural sciences department decreases the intention to protest the ADM system.

Table 2: OLS regression with dV voice

dV: Voice				
		B	SE	β
Step 1	Intercept	1.989	0.553	
	Gender (1=male)	−0.131	0.134	−.057
	Age	0.046	0.020	.160*
	Faculty	−0.450	0.165	−.159**
	Graduation status	0.065	0.145	.026
	Semester count	−0.014	0.026	−.037
	HDM DF	−0.074	0.084	−.054
	HDM PF	0.033	0.073	.027
Step 2	Intercept	3.825	0.605	
	Gender (1=male)	−0.176	0.126	−.076
	Age	0.029	0.019	.101
	Faculty	−0.310	0.157	−.110*
	Graduation status	−0.022	0.138	−.009
	Semester	−0.007	0.024	−.019
	HDM DF	0.111	0.085	.080
	HDM PF	−0.008	0.070	−.006
	ADM DF	−0.354	0.073	−.299**
	ADM PF	−0.152	0.062	−.142*
Note: *adj.* $R^2 = .162$, $\Delta R^2 = .112$. $^{*}p < .05$, $^{**}p < .01$				

Including the ADM-specific variables in the models shows that faculty membership still has a significant effect on *voice* ($\beta = -.11$), whereas the effect of age disappears. However, more importantly, the perceptions of both ADM fairness dimensions have a significant negative impact on *voice*, with the perception of distributive fairness having a stronger effect ($\beta = -.30$) than the perception of

[1] For that, we recoded the faculty (1 = Mathematics & Natural Sciences) and graduate degree variables (1 = Undergraduate) into dummy variables.
[2] Note that the human fairness perceptions are abbreviated as HDM DF for "HDM Distributive Fairness" and HDM PF for "HDM Procedural Fairness"; similarly, ADM fairness perceptions are abbreviated as ADM DF for "ADM Distributive Fairness" and ADM PF for "ADM Procedural Fairness."

procedural fairness ($\beta = -.14$). This means that perceiving the predicted outcome of ADM as unfair fuels the motivation to protest such a decision-making system. Perceiving the process itself as unfair and biased also adds to the intention to raise one's voice against ADM. Based on these results, we accept H3. Both procedural and distributive ADM fairness perceptions have a significant negative effect on voice.

The third regression model explains 14.5% of the variance in the dependent variable, *organizational reputation*. The largest share of variance is explained by the two ADM fairness perceptions ($\Delta R^2 = .13$). Again, faculty membership has a significant effect; it shows a positive impact on organizational reputation. However, the effect is not robust when adding the ADM-specific variables, as it is only significant at the p < .10 level. Nevertheless, students of mathematics or natural sciences tend to assign a better reputation to universities that adopt an AI-driven admission system ($\beta = .09$). In addition, distributive ADM fairness perceptions strongly predict ($\beta = .37$) a higher reputation of a university deploying AI-driven admission systems. Thus, if one considers admission through ADM as fair, the overall reputation of the university increases. Interestingly, the perception of procedural ADM fairness has no significant effect on reputation. For the university's reputation, whether the admission process is biased or unbiased seems irrelevant as long as the output of the decision is considered just. The findings, therefore, offer partial support for H4.

Table 3: OLS regression with dV organizational reputation

dV: Organizational Reputation				
		b	SE	β
Step 1	Intercept	3.343	0.388	
	Gender (1=male)	0.030	0.092	.019
	Age	−0.023	0.014	−.111
	Faculty	0.300	0.114	.151**
	Graduation status	−0.031	0.102	−.018
	Semester count	−0.006	0.018	−.023
	HDM DF	0.061	0.059	.062
	HDM PF	0.015	0.052	.018
Step 2	Intercept	2.054	0.422	
	Gender (1=male)	0.069	0.086	.043
	Age	−0.009	0.013	−.044
	Faculty	0.187	0.108	.094#
	Graduation status	0.039	0.095	.022
	Semester	−0.011	0.017	−.043
	HDM DF	−0.073	0.059	−.074
	HDM PF	0.040	0.049	.046
	ADM DF	0.314	0.050	.373**
	ADM PF	0.044	0.043	.058

Note: *adj.* $R^2 = .162$, $\Delta R^2 = .112$. #*p* < .10, **p* < .05, ***p* < .01

7 DISCUSSION AND CONCLUSION

This study shows that the use of ADM systems at universities has ambivalent consequences for the institution concerned. The expected consequences depend precisely on whether those who are affected perceive such systems as fair or biased. Our results suggest that perceptions of distributive fairness and procedural fairness negatively affect the intention to protest ADM usage for admission to public universities. They further show that procedural fairness perceptions have a negative effect on exit intentions and distributive fairness perceptions have a positive impact on organizational reputation. Thus, our study contributes to the growing literature on fairness perceptions of ADM in several ways.

First, concerning the comparison between ADM and HDM, we find that in terms of procedural and distributive fairness, people perceive AI-driven systems as fairer. This finding is in line with literature arguing that AI is perceived as fairer that humans in high-impact decisions [2]. Arguably, admission to public universities is an example of a high-impact decision for students. Looking more deeply, participants rated the procedural dimensions especially as fairer in the ADM scenario. It seems like HDM committees are perceived as more biased in their admission procedure. Yet students perceive ADM to be more objective and fairer. Furthermore, the distributive fairness dimension shows considerably large fairness differences. The output of the AI-driven admission system is considered to be much fairer than that of a human committee. Thus, regarding fairness perceptions, ADM is perceived as a fairer tool for students than HDM. Yet what are the subsequent effects of those perceptions?

H2–H4 address these questions. As predicted, all models show that at least one dimension of ADM fairness perception has a significant effect on behavioral intention or the university's reputation. Moreover, we see that the question of which fairness dimension is relevant seems to depend on the outcome variable. For example, in the case of reputation, only the distributive AI fairness perception seems to matter, whereas in the case of exit, only the procedural AI fairness dimensions has an effect. As outlined above, it is important to distinguish between different dimensions of fairness perceptions. Fairness is, thus, a multidimensional construct, as the organizational justice literature suggests [10], in that different dimensions account for different effects on different outcome variables. Future studies should pay closer attention to the multidimensionality of fairness perceptions and develop a more comprehensive framework for investigating perceived fairness.

We find that organizational reputation is only dependent on the assumed distributive AI fairness. Thus, if a university is concerned about its reputation among students, it should focus on the output of its AI-driven admission process. If applicants view university admissions by ADM as fair, it significantly increases the overall reputation of the university. In this case, it seems not to matter if and how people judge the process itself. Nevertheless, here lies a crucial problem: As students only receive information about their own classification and not the aggregate results, they may likely judge their own result as unfair if they are rejected. More research is needed to disentangle the individual and combined causal effects of distributive and procedural fairness on organizational reputation.

Concerning exit, however, we see that perceived procedural AI unfairness fuels students' intention to act by turning their back on the university and studying elsewhere. A possible explanation for this might be that students feel powerless because of the AI process, which leads to a decline in university spaces. Moreover, human procedural fairness perceptions fuel the intention to apply to a university that does not use an AI-based student admission system. The results raise a somewhat counterintuitive question about why some students might actually aim for a biased process, where they might influence the process and improve their chances of acceptance. However, these thoughts are somewhat speculative, and future research should address them.

Finally, we see that both distributive AI fairness perceptions and procedural AI fairness perceptions contribute to students' intention to raise their voice against an AI-driven student admission system by protesting. As unrest leaves a bad impression on the public, universities should take the satisfaction of their students into account. Ultimately, protest behavior on the part of students can result in tensions between the university and its students. This may have negative effects on student performance as well as on the reputation and operational efficiency of the university. However, voice is the more informative sign for a learning institution, while silent exit is the more dangerous threat. Thus, universities should provide sufficient opportunities for criticism and feedback when implementing ADM systems to prevent applicants from leaving immediately.

Notably, we also find that one particular student-specific variable matters in all three models—namely, faculty membership. Students of mathematics and natural sciences, overall, show less of an intention to protest the AI-driven admission system or to follow an exit strategy. They also tend to attribute a higher reputation to a university that uses advanced technologies. This leads us to conclude that students who arguably have more contact points with and greater knowledge of ADM technologies seem more open to the introduction of an AI-driven admission system. For universities that plan to introduce AI-driven systems, it is, therefore, important to address all stakeholders and consider that students from some faculties might react differently than others.

It is important to note that this study faced several limitations. First, we conducted this study in one German university. Therefore, the results are not representative of students of other universities (neither in Germany nor in other countries). Students of other universities (especially those universities that already use ADM) might have other attitudes and perceptions of AI technologies. Thus, future studies should take into account more universities and check if the effects found in this study are robust in different contexts. Second, we measured fairness perceptions as single items. Although many recent studies have used this measurement [2,8,26], we believe that a multi-item measurement of AI fairness is needed to account for more factors of the specific fairness dimensions.

In general, our study sheds light on several important points regarding fairness in the higher education context. First, it is important to differentiate between specific fairness dimensions. Fairness dimensions are judged specifically and have different effects on dependent variables. Second, if students think that they are receiving unfair treatment, they will intend to act against a university introducing ADM either in protest against the application or by turning their back on the university in general. Since universities receive funds according to the number of enrolled students, it is essential that they maintain or even increase their study body; thus, these perceptions are highly relevant. Furthermore, distributive fairness perceptions increase a university's reputation and, thus, give—in addition to the efficiency advantage—a potential incentive to implement AI-driven systems. Altogether, students' fairness perceptions are highly relevant when planning to introduce AI-driven systems in organizational processes.

8 ACKNOWLEDGEMENTS

Funded by the Volkswagen Foundation as part of the funding line "Artificial Intelligence and the Society of the Future".

REFERENCES

[1] Mustafa Acikkar and Mehmet Fatih Akay. 2009. Support vector machines for predicting the admission decision of a candidate to the School of Physical Education and Sports at Cukurova University. *Expert Syst. Appl.* 36, 3 (April 2009), 7228–7233. DOI:https://doi.org/10.1016/j.eswa.2008.09.007
[2] Theo Araujo, Natali Helberger, Sanne Kruikemeier, and Claes H. de Vreese. 2019. In AI We Trust? Usefulness, Fairness, and Risk Perceptions About Automated Decision-Making by Artificial Intelligence Within Media, Justice, and Health. In *ICA Conference*.
[3] Theo Araujo, Claes De Vreese, Natali Helberger, Sanne Kruikemeier, Julia van Weert, Nadine Bol, Daniel Oberski, Mykola Pechenizkiy, Gabi Schaap, and Linnet Taylor. 2018. *Automated Decision-Making Fairness in an AI-driven World: Public Perceptions, Hopes and Concerns*. Amsterdam.
[4] Mohsen Attaran, John Stark, and Derek Stotler. 2018. Opportunities and challenges for big data analytics in US higher education: A conceptual model for implementation. *Ind. High. Educ.* 32, 3 (2018), 169–182. DOI:https://doi.org/10.1177/0950422218770937
[5] Janine Baleis, Birte Keller, Christopher Starke, and Frank Marcinkowski. 2019. *Cognitive and Emotional Responses to Fairness in AI – A Systematic Review*. Düsseldorf. Retrieved from https://www.phil-fak.uni-duesseldorf.de/fileadmin/Redaktion/Institute/Sozialwissenschaften/Kommunikations-_und_Medienwissenschaft/KMW_I/Working_Paper/Baleis_et_al.__2019__Literatur_Review.pdf
[6] Solon Barocas and Andrew D. Selbst. 2016. Big Data's Disparate Impact. *104 Calif. Law Rev.* 671, 2016 (2016). DOI:https://doi.org/10.2139/ssrn.2477899
[7] Reuben Binns. 2018. Fairness in Machine Learning: Lessons from Political Philosophy. *Proc. Mach. Learn. Res.* 81, (December 2018), 1–11. Retrieved from http://arxiv.org/abs/1712.03586
[8] Reuben Binns, Max Van Kleek, Michael Veale, Ulrik Lyngs, Jun Zhao, and Nigel Shadbolt. 2018. "It's Reducing a Human Being to a Percentage." In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, 1–14. DOI:https://doi.org/10.1145/3173574.3173951
[9] Jeng-Fung Chen and Quang Hung Do. 2014. Training Neural Networks to Predict Student Academic Performance: A Comparison of Cuckoo Search and Gravitational Search Algorithms. *Int. J. Comput. Intell. Appl.* 13, 01 (March 2014), 1450005. DOI:https://doi.org/10.1142/S1469026814500059
[10] Jason A. Colquitt. 2001. On the dimensionality of organizational justice: A construct validation of a measure. *J. Appl. Psychol.* 86, 3 (June 2001), 386–400. DOI:https://doi.org/10.1037/0021-9010.86.3.386
[11] Ben Daniel. 2015. Big Data and analytics in higher education: Opportunities and challenges. *Br. J. Educ. Technol.* 46, 5 (2015), 904–920. DOI:https://doi.org/10.1111/bjet.12230
[12] Lina Dencik, Arne Hintz, Joanna Redden, and Emiliano Treré. 2019. Exploring Data Justice: Conceptions, Applications and Directions. *Information, Commun. Soc.* 22, 7 (June 2019), 873–881. DOI:https://doi.org/10.1080/1369118X.2019.1606268
[13] Morton Deutsch. 1975. Equity, Equality, and Need: What Determines Which Value Will Be Used as the Basis of Distributive Justice? *J. Soc. Issues* (1975). DOI:https://doi.org/10.1111/j.1540-4560.1975.tb01000.x
[14] Manuela Ekowo and Iris Palmer. 2016. The Promise and Peril of Predictive Analytics in Higher Education. A Landscape Analysis. October (2016).
[15] Federal Statistical Office. 2019. *Education, research and culture - Institutions of higher education*. Retrieved from https://www.destatis.de/EN/Themes/Society-Environment/Education-Research-Culture/Institutions-Higher-Education/_node.html#sprg266648
[16] Nina Grgic-Hlaca, Elissa M. Redmiles, Krishna P. Gummadi, and Adrian Weller.

2018. Human Perceptions of Fairness in Algorithmic Decision Making. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW '18*, 903–912. DOI:https://doi.org/10.1145/3178876.3186138
[17] Nina Grgic-Hlaca, Muhammad Bilal Zafar, Krishna P Gummadi, and Adrian Weller. 2018. Beyond Distributive Fairness in Algorithmic Decision. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*.
[18] Higher Education Compass. 2019. Degree Programmes. Retrieved December 5, 2019 from https://www.hochschulkompass.de/en/degree-programmes/application-admission/admissions-procedure.html
[19] Albert O. Hirschmann. 1970. *Exit, Voice and Loyalty. Responses to Decline in Firms, Organizations and States.* Harvard University Press, Cambridge.
[20] Anna Lauren Hoffmann. 2019. Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse. *Information, Commun. Soc.* 22, 7 (June 2019), 900–915. DOI:https://doi.org/10.1080/1369118X.2019.1573912
[21] Nicolas Huck. 2019. Large data sets and machine learning: Applications to statistical arbitrage. *Eur. J. Oper. Res.* 278, 1 (2019), 330–342. DOI:https://doi.org/10.1016/j.ejor.2019.04.013
[22] Ben Hutchinson and Margaret Mitchell. 2019. 50 Years of Test (Un)fairness. In *Proceedings of the Conference on Fairness, Accountability, and Transparency - FAT* '19*, 49–58. DOI:https://doi.org/10.1145/3287560.3287600
[23] Fei Jiang, Yong Jiang, Hui Zhi, Yi Dong, Hao Li, Sufeng Ma, Yilong Wang, Qiang Dong, Haipeng Shen, and Yongjun Wang. 2017. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc. Neurol.* 2, 4 (December 2017), 230–243. DOI:https://doi.org/10.1136/svn-2017-000101
[24] Birte Keller, Janine Baleis, Christopher Starke, and Frank Marcinkowski. 2019. *Machine Learning and Artificial Intelligence in Higher Education: A State-of-the-Art Report on the German University Landscape*. Düsseldorf. Retrieved from https://www.phil-fak.uni-duesseldorf.de/fileadmin/Redaktion/Institute/Sozialwissenschaften/Kommunikations-_und_Medienwissenschaft/KMW_I/Working_Paper/Keller_et_al.__2019__-_AI_in_Higher_Education.pdf
[25] Jakko Kemper and Daan Kolkman. 2018. Transparent to whom? No algorithmic accountability without a critical audience. *Information, Commun. Soc.* (June 2018), 1–16. DOI:https://doi.org/10.1080/1369118X.2018.1477967
[26] Min Kyung Lee. 2018. Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data Soc.* 5, 1 (June 2018), 205395171875668. DOI:https://doi.org/10.1177/2053951718756684
[27] Gerald S. Leventhal. 1980. What should be done with equity theory? New approaches to the study of fairness in social relationships. In *Social exchange: Advances in theory and research*, K.J. Gergen, M.S. Greenberg and R.H. Willis (eds.). Springer, Boston, 27–55.
[28] Jennifer M. Logg, Julia A. Minson, and Don A. Moore. 2019. Algorithm appreciation: People prefer algorithmic to human judgment. *Organ. Behav. Hum. Decis. Process.* (2019). DOI:https://doi.org/10.1016/j.obhdp.2018.12.005
[29] Frank Marcinkowski and Christopher Starke. 2019. Wann ist Künstliche Intelligenz (un-)fair? Ein sozialwissenschaftliches Konzept von KI-Fairness [When is Artificial Intelligence (Un-)Fair? A Social Science Concept of AI-Fairness. In *Politik in der digitalen Gesellschaft. Zentrale Problemfelder und Forschungsperspektiven [Politics in the Digital Society. Important Problems and Research Perspectives]*, Jeanette Hofmann, Norbert Kersting, Claudia Ritzi and Wolf J. Schünemann (eds.). transcript, Bielefeld, 269–288. DOI:https://doi.org/10.14361/9783839448649
[30] Paul K. McClure. 2018. "You're Fired," Says the Robot. *Soc. Sci. Comput. Rev.* 36, 2 (April 2018), 139–156. DOI:https://doi.org/10.1177/0894439317698637
[31] Neema Mduma, Khamisi Kalegele, and Dina Machuve. 2019. A Survey of Machine Learning Approaches and Techniques for Student Dropout Prediction. *Data Sci. J.* 18, 1 (April 2019). DOI:https://doi.org/10.5334/dsj-2019-014
[32] Leonard J Ponzi, Charles J Fombrun, and Naomi A Gardberg. 2011. RepTrak™ Pulse: Conceptualizing and Validating a Short-Form Measure of Corporate Reputation. *Corp. Reput. Rev.* 14, 1 (May 2011), 15–35. DOI:https://doi.org/10.1057/crr.2011.5
[33] Donghee Shin and Yong Jin Park. 2019. Role of fairness, accountability, and transparency in algorithmic affordance. *Comput. Human Behav.* 98, (September 2019), 277–284. DOI:https://doi.org/10.1016/j.chb.2019.04.019
[34] Megha Srivastava, Hoda Heidari, and Andreas Krause. 2019. Mathematical Notions vs. Human Perception of Fairness. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining - KDD '19*, 2459–2468. DOI:https://doi.org/10.1145/3292500.3330664
[35] Linnet Taylor. 2017. What is data justice? The case for connecting digital rights and freedoms globally. *Big Data Soc.* 4, 2 (December 2017), 205395171773633. DOI:https://doi.org/10.1177/2053951717736335
[36] Eric J. Topol. 2019. High-performance medicine: the convergence of human and artificial intelligence. *Nat. Med.* 25, 1 (January 2019), 44–56. DOI:https://doi.org/10.1038/s41591-018-0300-7
[37] Sahil Verma and Julia Rubin. 2018. Fairness definitions explained. In *Proceedings of the International Workshop on Software Fairness - FairWare '18*, 1–7. DOI:https://doi.org/10.1145/3194770.3194776
[38] Bernd W. Wirtz, Jan C. Weyerer, and Carolin Geyer. 2019. Artificial Intelligence and the Public Sector—Applications and Challenges. *Int. J. Public Adm.* 42, 7 (May 2019), 596–615. DOI:https://doi.org/10.1080/01900692.2018.1498103
[39] Tan Fee Yean and Ab Aziz Yusof. 2016. Organizational Justice: A Conceptual Discussion. *Procedia - Soc. Behav. Sci.* 219, (May 2016), 798–803. DOI:https://doi.org/10.1016/j.sbspro.2016.05.082
[40] Olaf Zawacki-Richter, Victoria I. Marín, Melissa Bond, and Franziska Gouverneur. 2019. Systematic review of research on artificial intelligence applications in higher education – where are the educators? *Int. J. Educ. Technol. High. Educ.* 16, 39 (December 2019). DOI:https://doi.org/10.1186/s41239-019-0171-0
[41] Jonathan Zittrain. 2019. The Hidden Costs of Automated Thinking. *The New Yorker*. Retrieved July 29, 2019 from https://www.newyorker.com/tech/annals-of-technology/the-hidden-costs-of-automated-thinking

APPENDIX

Translated version of the vignette utilized in the questionnaire for introducing the general functionality of ADM systems to participants:

Currently, there is a lot of public talk about "artificial intelligence" (AI). This refers to computer applications that automatically evaluate digital data. For AI, the evaluation of large amounts of data represents a learning process from which it continuously processes new information and, thus, recognizes increasingly precise patterns over time. On the basis of this analysis, facts can be established and future developments can be forecast. Systems with artificial intelligence can propose recommendations for action to humans or make autonomous decisions and execute them directly.

Translated version of the vignette utilized in the questionnaire for introducing the use of ADM systems for higher education admissions to participants:

Let us now take a closer look at a possible application of artificial intelligence in higher education—namely, admission to a study course.

Committees comprising members who advise on the suitability of applicants and decide on admission (so-called admission or selection committees) usually allocate students to degree programs with restricted admissions.

Recently, artificial intelligence technologies that can take on this task have become available. Such AI applications analyze the available data to determine who is particularly suitable for a particular subject. On this basis, the system issues an approval or rejection recommendation.