Explainable Artificial Intelligence for Academic Performance Prediction. An Experimental Study on the Impact of Accuracy and Simplicity of Decision Trees on Causability and Fairness Perceptions

Marco Lünich
marco.luenich@hhu.de
Heinrich Heine University Düsseldorf
Düsseldorf, Germany

Birte Keller
birte.kellerh@hhu.de
Heinrich Heine University Düsseldorf
Düsseldorf, Germany

ABSTRACT

The rising adoption of learning analytics and academic performance prediction technologies in higher education highlights the urgent need for transparency and explainability. This demand, rooted in ethical concerns and fairness considerations, converges with Explainable Artificial Intelligence (XAI) principles. Despite the recognized importance of transparency and fairness in learning analytics, empirical studies examining student fairness perceptions, particularly within academic performance prediction, remain limited. We conducted a pre-registered factorial survey experiment involving 1,047 German students to investigate how decision tree features (simplicity and accuracy) influence perceived distributive and informational fairness, mediated by causability (i.e., the self-assessed understandability of a machine learning model's cause-effect linkages). Additionally, we examined the moderating role of institutional trust in these relationships. Our results indicate that decision tree simplicity positively affects fairness perceptions, mediated by causability. In contrast, prediction accuracy neither directly nor indirectly influences these perceptions. Even if the hypothesized effects of interest are either minor or non-existent, results show that the medium positive effect of causability on the distributive fairness assessment depends on institutional trust. These findings substantially impact the crafting of transparent machine learning models in educational settings. We discuss important implications for fairness and transparency in implementing academic performance prediction systems.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in visualization**; **User studies**; **Laboratory experiments**; • **Computing methodologies** → **Artificial intelligence**.

KEYWORDS

Explainable AI, Trustworthy AI, ML Fairness, Academic Performance Prediction, Experiment, Causability

FAccT '24, June 03–06, 2024, Rio de Janeiro, Brazil

ACM ISBN 979-8-4007-0450-5/24/06
https://doi.org/10.1145/3630106.3658953

ACM Reference Format:
Marco Lünich and Birte Keller. 2024. Explainable Artificial Intelligence for Academic Performance Prediction. An Experimental Study on the Impact of Accuracy and Simplicity of Decision Trees on Causability and Fairness Perceptions. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24), June 03–06, 2024, Rio de Janeiro, Brazil.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3630106.3658953

1 INTRODUCTION

Higher education institutions are exploring academic performance prediction (APP) using machine learning (ML) in learning analytics (LA). Ensuring transparency and explainability is crucial, mandated by EU regulations and ethical guidelines, as well as student demand [43]. This urgency is rooted in social fairness issues, where Explainable Artificial Intelligence (XAI) is relevant [1, 34, 36, 46]. Though transparency is emphasized in educational policies [14], "the number of empirical user studies examining ethical considerations, such as transparency in AI, is relatively low and often focused on LA in general rather than specific LA systems" [19]. Our research thus focuses on students' perceptions of APP fairness. Using a factorial survey and a pre-registered experiment with 1,047 German students, we evaluated the impact of decision tree features—model simplicity and accuracy—on perceived fairness. We also examined the mediating role of causability, i.e., the self-assessed understandability of the cause-effect relationships of a ML model [40, 78] in these relationships and assessed if institutional trust moderates these effects.

2 ACADEMIC PERFORMANCE PREDICTION AS A FORM OF AI-BASED LEARNING ANALYTICS

Within the last few years, many LA applications have been developed and adopted worldwide to support the work of students, lecturers, and administrators alike [16, 84]. Some of those applications can be used to reach several goals at once. For instance, with the implementation of APP systems, higher education institutions (HEI) aim to improve student success and reach higher equality in retention rates [5, 67]. APP systems are being used to predict students' performance based on large amounts of data–mainly historical performance data, but in some cases also sociodemographic data [2, 27]–with the help of ML [4]. The actual prediction can vary from the prediction of students' grades [6, 20] over the likelihood of a successful study completion to the prediction of a potential

dropout [12, 64]. Moreover, when APP is used to give individualized feedback or to distribute support measures, APP can further help especially those students who have often been disadvantaged before, by providing them with a more tailored educational experience that ensures that each student's unique needs and potential are recognized and addressed [60, 67].

2.1 The Issues of Discrimination and Fairness Perceptions

However, discrimination and fairness are critical concerns in applying LA and APP [9]. Discrimination can arise from flawed algorithmic design, biased data, or actions based on AI predictions [49, 63]. These issues can perpetuate societal biases [29, 58] and influence human decision-making [49]. Efforts exist to enhance the fairness of LA systems [26, 42, 61]. *Perceived fairness* issues are equally important [46]. APP applications are sociotechnical systems; thus, stakeholder perceptions, particularly students', are crucial [38, 51, 55]. Ignoring students' perceptions can negatively impact course satisfaction [72], diminished reputation for, or even outright rejection of HEIs [59], or provoke public protest [28]. Hence, stakeholder involvement in APP design is essential [46, 70]. Our study focuses on students' perceptions of distributive and informational fairness. Distributive fairness involves outcome evaluations [46], while informational fairness relates to decision-making transparency [22]. Given ethical guidelines advocating transparency in AI [13, 14, 43], and students' lack of awareness about data use [44], transparency is vital for perceived fairness [9, 46].

2.2 On the Value of Explainable AI for APP Fairness

To achieve fairness and trust in LA, recent research has turned towards XAI methods [91]. According to Gunning et al., "the purpose of an explainable AI (XAI) system is to make its behavior more intelligible to humans by providing explanations. (...) The XAI system should be able to explain its capabilities and understandings; explain what it has done, what it is doing now, and what will happen next; and disclose the salient information that it is acting on" [35]. Adadi and Berrada identify four reasons for the necessity of XAI: justification, control, improvement of AI systems, and knowledge production [1]. XAI is deemed critical for individuals to understand and verify decisions [10] and is considered a prerequisite for fairness [34]. Empirical evidence suggests that explanations improve perceived fairness and trust in AI systems [8, 25, 78, 79, 82]. Specifically, explanations have shown to affect informational fairness [76, 89]. However, the impact varies depending on the dimensions of fairness and types of explanations used [11, 25, 75, 77, 79]. Despite the demand for transparency, particularly from students [68, 80, 86], there is still a gap in understanding user prerequisites, needs and, expectations [18, 30, 47]. In the context of LA, decision trees are often used for XAI in APP [41]. These decision trees are trained on historical data, such as past exam results, to predict future academic outcomes [36, 50]. By making explanatory factors transparent, biases can be identified more quickly, contributing to the development of fairer algorithms [54]. However, the complexity and accuracy of decision trees can vary [50].

2.3 Simplicity and Fairness Perceptions

As ML models grow in complexity, there is a risk they may become too intricate for human comprehension. Balancing complexity and simplicity is essential for explainability, although no definitive standards exist [74]. Cognitive limitations further complicate the issue; for instance, young adults can process only three to five stimuli at the same time [21]. Empirical studies reveal a nuanced relationship between informational fairness and the amount of explanation provided [23, 48, 76]. However, the theory of explanatory coherence suggests that simpler explanations are generally preferred [62, 83]. This preference for simplicity has been empirically supported and observed in various applications, including health symptom checks and mathematical fairness notions [69, 81, 87]. Yurrita et al.'s qualitative study indicated that too much information or complexity could be counterproductive, especially for those with limited AI literacy [89]. Hence, the simplicity of an APP decision tree is critical to ensure equitable understanding and to prevent potential discrimination, such as favoring students with prior computer science knowledge. Based on these considerations, we formulate the first hypothesis:

Hypothesis 1 (H1). Simpler decision trees lead to higher perceived informational fairness.

2.4 Accuracy and Fairness Perceptions

Accuracy is crucial for adopting and trusting APP systems. Low accuracy hinders adoption and impacts the system's fairness [37, 49]. Literature on algorithmic aversion indicates that observed errors in AI systems reduce people's confidence in them [24, 57]. Accuracy strongly predicts intention to follow AI recommendations, even more so than clarity of origin [32]. It also positively affects trust in AI [65, 66, 88]. However, Conijn et al. found no effect of accuracy explanations on student motivation in an essay grading context [19]. Given the significance of accuracy and the inherent trade-offs in system design–since principles like transparency, explainability, and accuracy cannot be simultaneously maximized within a single ML model or XAI approach [3]–we formulate the second hypothesis:

Hypothesis 2 (H2). Decision trees with a higher accuracy lead to higher perceived distributive fairness.

2.5 Fairness and Causability

Simplicity in APP systems aims to improve informational fairness but does not guarantee understandability for all students. If explanations are accessible only to a subset of students, unfairness ensues [10]. In this regard, the concept of *causability*, introduced by Holzinger et al., assesses the quality of explanations from the user's perspective [40]. Unlike explainability, which focuses on the system's capability to elucidate its functions [35], causability is user-centric and measures a person's understanding of an explanation. Holzinger et al. operationalized this with the *system causability scale* [39]. Shin's study on AI journalism supports causability's role as an antecedent to explainability and its influence on perceived fairness and trust [78]. Based on these insights, we formulate the third and fourth hypotheses:

Hypothesis 3 (H3). The relationship between the simplicity of decision trees and perceived informational fairness is mediated by causability, such that simpler decision trees lead to greater causability, which in turn leads to increased perceived informational fairness.

Hypothesis 4 (H4). The relationship between the accuracy of decision trees and perceived distributive fairness is mediated by causability, such that decision trees with a higher accuracy lead to greater causability, which in turn leads to increased perceived distributive fairness.

2.6 Fairness and Institutional Trust

HEIs, as the stewards of APP systems, carry the ethical responsibility to safeguard students from the consequences of unfair decisions [45]. Failure in this regard risks eroding institutional trust, as seen in cases involving automated assessment discrimination [28] and unethical data use [45]. However, trust in HEIs significantly influences students' willingness to disclose data for LA [53, 80]. In the sociotechnical landscape of APP, trust extends beyond the technology to include the organizations and individuals that deploy it [56, 85]. Trust also acts as a complexity-reducing mechanism in uncertain situations [52]. Therefore, students' inherent trust in their HEIs could potentially mitigate the need to fully comprehend APP's intricacies, assuming the institution is perceived as ethical and trustworthy [80, 90]. Based on these considerations, we formulate the fifth and sixth hypotheses:

Hypothesis 5 (H5). The indirect effect of the simplicity of decision trees on perceived informational fairness through causability is moderated by institutional trust. Specifically, simpler decision trees lead to greater causability, which in turn results in increased perceived informational fairness. However, the strength of this mediated relationship is contingent upon the level of institutional trust. The indirect effect is weaker at higher levels of institutional trust, as individuals with high institutional trust have a higher perception of informational fairness even when simplicity and accuracy are low, making the role of causability as a mediator less influential in these cases.

Hypothesis 6 (H6). The indirect effect of the accuracy of decision trees on perceived distributive fairness through causability is moderated by institutional trust. Specifically, simpler decision trees lead to greater causability, which in turn results in increased perceived distributive fairness. However, the strength of this mediated relationship is contingent upon the level of institutional trust. The indirect effect is weaker at higher levels of institutional trust, as individuals with high institutional trust have a higher perception of distributive fairness even when simplicity and accuracy are low, making the role of causability as a mediator less influential in these cases.

Our primary focus is on assessing the impact of accuracy on distributive fairness and simplicity on informational fairness, with attention to the roles of causability and institutional trust. Our conceptual moderated mediation model is illustrated in Figure 1. In this model, two paths are not hypothesized to have direct effects. Subsequently, these paths will be freely estimated in our structural regression analysis, allowing for data-driven insights as this study aims to elucidate how simplicity and accuracy in decision trees affect fairness perceptions. Consequently, we pose two research questions:

RQ1. To what extent do simple decision trees affect perceived distributive fairness?

RQ2. To what extent do accurate decision trees affect perceived informational fairness?

All hypotheses and research questions were pre-registered via OSF.

3 METHOD

In a 2x2 between-subjects experimental design, we examine the influence of decision tree simplicity and accuracy on students' causability and subsequent fairness perceptions. We also evaluate the moderating role of institutional trust on the causability-fairness relationship through a moderated mediation model. Data analysis was carried out using the statistical program R version 4.3.1 (2023-06-16 ucrt) using structural equation modeling with the package *lavaan* [73]. All estimated models utilize bootstrapping with 5000 bootstraps. Bootstrap intervals for parameter estimates were produced using the adjusted bootstrap percentile method with bias correction.

The sample size for our study was determined through an a priori power analysis conducted in R, aiming for a power level of .75 with an anticipated sample size of around 1000 participants. This increased to approximately .8 with 1100 participants. We targeted a small effect size difference of .15 for the moderation of a mediation effect at a conventional alpha error probability of .05. Despite the risk of a 25% Type II error, we deemed this power level acceptable for our specific research context, balancing data collection feasibility with the robustness of our results. As the questionnaire needed to be accessed via a non-mobile device to display the stimulus correctly, the maximum number of respondents available via the panel provider was approximately 1100. Participants were recruited from the online panel Talk Online Data Collection AG, were 18 or older, and enrolled in higher education. The Ethical Review Board of the Faculty of Philosophy of the Heinrich Heine University Düsseldorf, Germany, approved the study.

All in all, n = 1047 students completed the survey. The average age of students was 25.46 (SD = 5.6). Altogether, 572 (54.7%) students identified as women, 462 (44.2%) as men, and 11 (1.1%) did not identify strictly as male or female, indicating 'diverse'. Of all students, 393 (37.5%) indicated studying a STEM subject and 606 (57.9%) indicated studying a non-STEM subject with 48 (4.6%) students indicating to study something else which could not be assigned to either STEM or non-STEM subjects. Regarding the question of which degree students are currently pursuing in their core subject, 636 (60.7%) students report pursuing a bachelor's degree, 246 (23.5%) report pursuing a master's degree, 129 (12.3%) report pursuing state examination, and 36 (3.4%) pursue a doctorate.

3.1 Procedure and Survey Design

First, respondents were briefed on the study's objectives, questionnaire duration, and data protection measures. After giving informed consent, they confirmed current enrollment in a HEI. Sociodemographic data and institutional trust levels were then collected. A brief overview of AI and Academic Performance Prediction APP was provided before randomly presenting one of four decision tree stimuli, varying in complexity and accuracy. A treatment check

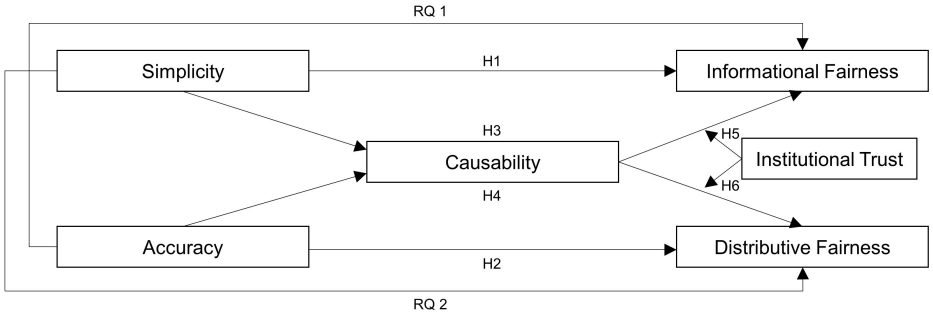

Figure 1: Conceptual Moderated Mediation Model

and questions about the decision tree's causability followed. Participants then evaluated the tree's distributive and informational fairness before answering additional academic-related questions. Finally, they were debriefed, redirected to the panel provider, and compensated. The average time to complete the questionnaire was 6.06 minutes (SD = 2.55).

3.2 Measurements

Independent Variable (IV). Simplicity

As stimuli, students viewed one of four decision trees detailing factors affecting APP. The trees varied in simplicity, serving as an independent variable. The simpler tree had two decision levels leading to an outcome prediction, while the more complex one had up to five decision rules. These factors were based on a real APP system in development [27] and excluded potentially discriminatory data. Tailored to computer science and social sciences students, the factors were abstracted for cross-field comparability, allowing identification for respondents across subjects.

IV. Accuracy

Regarding accuracy, the second independent variable, rates differed between a higher accuracy of 95% and a lower accuracy of 65%. This information was displayed conspicuously below the decision tree. These rates were chosen to capture performance variance. In the 2x2 between-subject design, students viewed a decision tree that was either simple or complex and had an accuracy of either 65% or 95%. For visual reference, see Figures 2 and 3.

DV. Informational Fairness.

As a dependent variable (DV), we focused on students' perceptions of fairness. Thus, information fairness was measured with several items in the first step. To achieve good factorial validity (Cronbach's α = 0.80; AVE = 0.57), we decided to choose the following three items for the latent variable of our structural equation model: "The reasons for the prediction are understandable."; "The explanation of the AI-based performance prediction procedure is comprehensive."; "The explanation of the prediction is coherent." The first item is self-developed, while the other two were adapted from Schoeffer et al. [77]. All variables used for the study were measured on a five-point Likert scale ranging from 1 = "strongly disagree" to 5 = "strongly agree". Respondents also had the option of expressing no preference ("don't know"; except in the case of institutional trust).[1]

DV. Distributive Fairness.

Students' perceived distributive justice was measured using three commonly used items developed by Colquitt and Rodell [17] and adapted to the APP context. Students rated whether they agreed with the following statements: "The prediction of performance for students by AI is fair."; "Everyone gets what he/she deserves."; "No one is unduly disadvantaged by student performance prediction by AI." All three items also show good factorial validity (Cronbach's α = 0.83; AVE = 0.62).

With the two dependent variables, informational fairness and distributive fairness, demonstrating good convergent validity, the question arises as to what extent they are distinct constructs, addressing the issue of discriminant validity. A high correlation is observable in a model where both constructs are estimated as individual latent factors, r = 0.68. However, discriminant validity assessment does not only focus on the observed correlation between the constructs. By applying the Fornell-Larcker criterion [31], which stipulates that the squared correlation between the constructs must be less than the individual AVE of each factor, results suggest that there is discriminant validity, r^2 = 0.46.

Mediator. Causability

To assess students' understanding of the APP presented, namely causability, we used the *System Causability Scale* from Holzinger et al. [39]. Again, however, we had to exclude items to improve factorial validity. Respondents were asked to indicate how much they agreed with the following statements regarding the explanations in APP's decision tree: "I understand the explanations in the context of my studies."; "The explanations help me understand the criteria of the prediction."; "I am able to understand the explanations with my prior knowledge." Previously, it was explained that the term "explanations" refers to the decision process shown earlier, which shows at the end whether a dropout is predicted by the AI system or not. The items show good factorial validity (Cronbach's α = 0.78; AVE = 0.54).

Moderator. Institutional Trust.

[1]To compare groups, it is crucial to ensure that the latent constructs measured using multiple items demonstrate measurement invariance. We assessed measurement invariance across the groups within the experimental conditions for the latent constructs. Following the criterion suggested by Cheung and Rensvold [15], deviations in the CFI did not exceed .01, suggesting that the measurements maintain consistent psychometric properties and factor structure across different conditions.

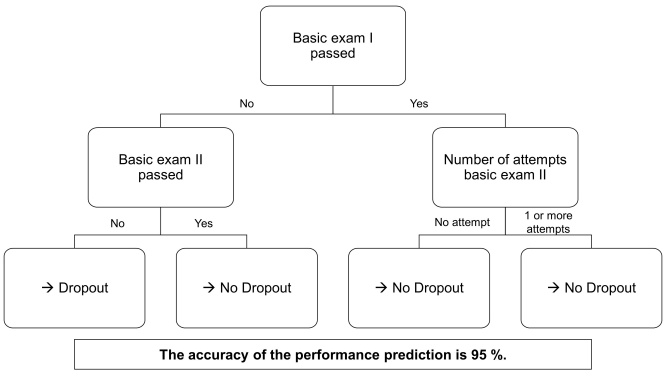

Figure 2: Simple Decision Tree for Academic Performance Prediction with High Accuracy

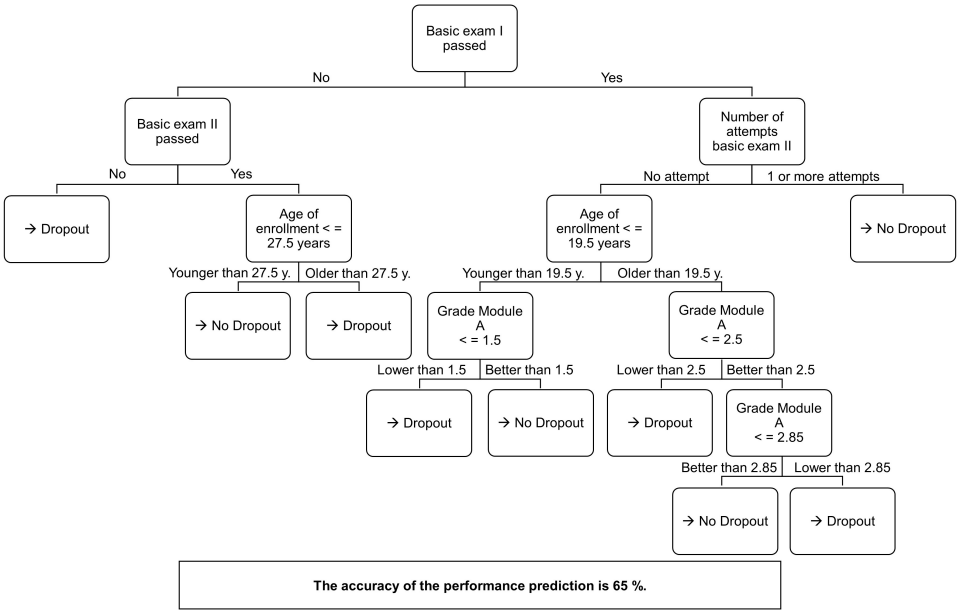

Figure 3: Complex Decision Tree for Academic Performance Prediction with Lower Accuracy

Finally, institutional trust was measured regarding the HEI at which students are enrolled. We used and slightly adapted four items developed by Gosh et al. [33] and previously validated by Li et al. [53]. The students were asked to rate the extent to which they agreed with the following sentences: "Since I cannot personally supervise all of my university's activities, I rely on the university staff to do their jobs properly."; "I believe that my university is a credible organization."; "I feel that I can rely on my university.". However, to improve factorial validity, we excluded the reverse coded item ("In general, I do not have confidence in my university."). After doing so, the items show good factorial validity (Cronbach's α = 0.75; AVE = 0.53).

3.3 Treatment Check

Two treatment check items assessed participants' perceptions of the manipulated conditions. Participants responded on a five-point scale, ranging from 1 (strongly disagree) to 5 (strongly agree). The first item examined the perception of decision tree simplicity: *If you think about the decision tree just shown: To what extent do you agree or disagree with the following statements? The decision tree of the performance prediction is straightforward.* The results indicated a significant difference among the four conditions ($F(3, 577.9) = 42.8$, $p < 0.001$). Using a Games-Howell post hoc test, the conditions with a less simple decision tree and high accuracy (M = 3.40; SD = 1.11) and lower accuracy, respectively (M = 3.45; SD = 1.16), were found to differ significantly from the conditions with a simpler decision tree and high accuracy (M = 4.17; SD = 0.93) and lower accuracy, respectively (M = 4.14; SD = 0.96), confirming that respondents recognized the extent to which the displayed simplicity of the decision tree varied. The second item assessed the perception of decision tree accuracy: *The performance prediction decision tree shows high accuracy.* The results indicated a significant difference

among the four conditions (F(3, 1043) = 20.34, p < 0.001). Using a Games-Howell post hoc test, the conditions with high accuracy and a less simple decision tree (M = 3.31; SD = 1.05) and simpler decision tree, respectively (M = 3.36; SD = 1.06), were found to differ significantly from the conditions with low accuracy and a less simple decision tree and (M = 2.77; SD = 1.08) and simpler decision tree, respectively (M = 2.89; SD = 1.08), confirming that respondents recognized the extent to which the displayed accuracy of the decision tree varied.

4 RESULTS

4.1 Main Effects Model

We first address a main effects model to test H1 and H2, reporting the direct effects of the exogenous variables simplicity and accuracy that were manipulated in the stimulus as IVs on the DVs informational and distributive fairness. For the sake of completeness, in this analysis, we estimated a model that also includes the interaction of simplicity and accuracy. The structural regression model shows good fit ($\chi^2(20) = 68.89$, $p = < 0.001$; *RMSEA* = 0.05, *90% CI* [0.04, 0.06]; *TLI* = 0.97).

Regarding the simplicity of the decision tree, results suggest that there is a small positive and significant effect of the simplicity of the decision tree on perceptions of informational fairness; the simpler the decision tree, the greater the perceived informational fairness, B = 0.32, *SE* = 0.10, *95% CI* (0.14, 0.51), $p = < 0.001$, $\beta = 0.16$. Accordingly, H1 is accepted. Additionally, there is a small positive and significant effect of simplicity on perceived distributive fairness; the simpler the decision tree, the greater the perceived distributive fairness, B = 0.31, *SE* = 0.12, 95% CI (0.07, 0.52), $p = 0.008$, $\beta = 0.12$.

Regarding the accuracy of the decision tree, there was neither an effect on perceptions of informational fairness, B = 0.06, *SE* = 0.10, *95% CI* (-0.14, 0.25), $p = 0.553$, $\beta = 0.03$, nor on on perceived distributive fairness, B = -0.01, *SE* = 0.12, 95% CI (-0.24, 0.22), $p = 0.913$, $\beta = -0.01$. Accordingly, H2 is rejected.

Lastly, there was neither an interaction effect of the IVs on perceptions of informational fairness, B = -0.14, *SE* = 0.14, *95%* CI (-0.41, 0.14), $p = 0.323$, $\beta = -0.06$, nor on perceived distributive fairness, B = -0.12, *SE* = 0.16, *95% CI* (-0.43, 0.21), $p = 0.470$, $\beta = -0.04$.

4.2 Mediation Model

Second, to test H3 and H4, we estimated a mediation model that integrates the mediator, causability, to understand how the IVs affect the DVs through an indirect pathway. Specifically, the model not only tests whether simplicity and accuracy directly affect both informational and distributive fairness but also influence them indirectly via causability. The structural regression model shows good fit ($\chi^2(36) = 99.30$, $p = < 0.001$; *RMSEA* = 0.04, *90% CI* [0.03, 0.05]; *TLI* = 0.98).

Regarding the effects of the IVs on the mediator, causability (i.e., the first path of the indirect effect), we first examine the impacts of simplicity and accuracy. For the simplicity of the decision tree, there was positive effect on causability; the simpler the decision tree, the greater the causability. The effect is small and significant, B = 0.16, *SE* = 0.07, *95% CI* (0.03, 0.29), $p = 0.016$, $\beta = 0.08$. For the accuracy of the decision tree, there was no significant effect on causability, B = 0.04, *SE* = 0.07, *95% CI* (-0.10, 0.17), $p = 0.552$, $\beta = 0.02$.

Second, we assess the effects of the mediator, causability, on the DVs informational fairness and distributive fairness (i.e., the second path of the indirect effect). First, there was positive effect of causability on informational fairness; the higher the causability, the greater the perceived informational fairness. The effect is strong and significant, B = 0.72, *SE* = 0.05, *95% CI* (0.62, 0.82), $p = < 0.001$, $\beta = 0.69$. Second, there was positive effect of causability on distributive fairness; the higher the causability, the greater the perceived distributive fairness. The effect is strong and significant, too, B = 0.55, *SE* = 0.05, *95% CI* (0.45, 0.66), $p = < 0.001$, $\beta = 0.43$.

Altogether, the results showed that there was a significant total effect between simplicity and informational fairness, B = 0.25, *SE* = 0.07, *95% CI* (0.11, 0.39), $p = < 0.001$, $\beta = 0.13$. Controlling for causability, the direct effect of simplicity on informational fairness remained significant, indicating that the effect is partially mediated, B = 0.14, *SE* = 0.06, *95% CI* (0.01, 0.25), $p = 0.025$, $\beta = 0.07$. The indirect effect was significant supporting the presence of a mediation effect, B = 0.12, *SE* = 0.05, *95% CI* (0.02, 0.22), $p = 0.019$, $\beta = 0.06$. Accordingly, H3 is accepted.

Furthermore, the results showed that there was a significant total effect between simplicity and distributive fairness, B = 0.25, *SE* = 0.08, *95% CI* (0.08, 0.41), $p = 0.004$, $\beta = 0.10$. Controlling for causability, the direct effect of simplicity on distributive fairness remained significant, indicating that the effect is partially mediated, B = 0.16, *SE* = 0.08, *95% CI* (0.01, 0.32), $p = 0.049$, $\beta = 0.06$. The indirect effect was significant supporting the presence of a mediation effect, B = 0.09, *SE* = 0.04, *95% CI* (0.02, 0.17), $p = 0.023$, $\beta = 0.04$.

Contrarily, the results showed that there was no significant total effect between accuracy and distributive fairness, B = -0.07, *SE* = 0.08, *95% CI* (-0.24, 0.09), $p = 0.389$, $\beta = -0.03$. Controlling for causability, the direct effect of accuracy on distributive fairness was non-significant, B = -0.09, *SE* = 0.08, *95% CI* (-0.25, 0.06), $p = 0.230$, $\beta = -0.04$. The indirect effect was non-significant, too, B = 0.02, *SE* = 0.04, *95% CI* (-0.06, 0.09), $p = 0.551$, $\beta = 0.01$. Accordingly, H4 is rejected.

Likewise, the results showed that there was no significant total effect between accuracy and informational fairness, B = -0.01, *SE* = 0.07, *95% CI* (-0.15, 0.12), $p = 0.883$, $\beta = -0.01$. Controlling for causability, the direct effect of accuracy on informational fairness was non-significant, B = -0.04, *SE* = 0.06, *95%* CI (-0.16, 0.07), $p = 0.510$, $\beta = -0.02$. The indirect effect was non-significant, too, B = 0.03, *SE* = 0.05, *95% CI* (-0.07, 0.12), $p = 0.552$, $\beta = 0.01$.

4.3 Moderated Mediation Model

Lastly, to test H5 and H6, we estimated a moderated mediation model that integrates the moderator, institutional trust, to test whether the effect of causability on informational fairness and distributive fairness as part of the indirect effect of the IV on the DV via the mediator depends on the extent of students' institutional trust. We first report a model with parameter estimates that treat the moderator as a continous latent variable. Second, we compare the model across three groups to illustrate the changes in the effect at different levels of the moderator: a) a group of students whose institutional trust scores are at least one standard deviation (SD)

below the mean, b) a group whose institutional trust is within one SD of the mean, and c) a group whose institutional trust is one SD above the mean.

The first estimated model, which treats the moderator as a continuous latent variable, suggests a poor fit ($\chi^2(213) = 1851.31$, $p = < 0.001$; *RMSEA* = 0.09, *90% CI* [0.08, 0.09]; *TLI* = 0.81). This poor fit arises from the introduction of the interaction terms, which multiply both the amount and magnitude of covariances between variables that are assumed to be independent. When combined with the sample size, this increase in covariance elevates the chi-square value, a measure of goodness of fit, possibly leading to rejection of the model under conventional thresholds. However, the model does reach a plausible solution, and an examination of the parameter estimates indicates no major deviation compared to the previously estimated models.

The parameter estimate for the first interaction effect suggests no effect of institutional trust on the relationship between the causability and informational fairness, B = -0.04, -0.03, *SE* = 0.05, *95% CI* (-0.12, 0.09), $p = 0.591$, $\beta = -0.02$.

Figure 4 below shows the parameter estimate for the effect of causability on informational fairness for students at least one SD below the mean of institutional trust, within one SD of the mean, and at least one SD above the mean. As there is no moderation effect, we reject H5.

The parameter estimate for second interaction effect suggests a small negative and significant effect of institutional trust on the relationship between the causability and distributive fairness, B = -0.11, *SE* = 0.06, *95% CI* (-0.21, 0.01), $p = 0.054$, $\beta = -0.07$.

Figure 5 below shows the parameter estimate for the effect of causability on distributive fairness for students at least one SD below the mean of institutional trust, within one SD of the mean, and at least one SD above the mean. While our confidence intervals for the individual parameter estimates in the figure do overlap, suggesting uncertainty in the distinctions between some groups, our more direct tests of differences between specific groups show significant contrasts. For instance, the difference of the regression parameter 'Distributive Fairness ~ Causability' between 'More than 1 SD below Mean' and 'More than 1 SD above Mean' is statistically significant, B = 0.42, *SE* = 0.21, *95% CI* (0.04, 0.86), $p = 0.046$, $\Delta\beta = 0.29$. This suggests that even though the confidence bands for these groups might overlap when calculated and visualized individually, the statistical evidence points towards a difference in their actual parameter estimates. However, as there was overall no total effect of accuracy on distributive fairness and no indirect effect via the causability, the different indirect effects for the first group, B = -0.00, *SE* = 0.14, *95% CI* (-0.26, 0.28), $p = 0.981$, $\beta = -0.00$, the second group, B = 0.01, *SE* = 0.05, *95% CI* (-0.08, 0.09), $p = 0.831$, $\beta = 0.00$, and the third group, B = 0.09, *SE* = 0.07, *95% CI* (-0.04, 0.26), $p = 0.236$, $\beta = 0.03$, are too small to reach significance and suggest practically no effect for fairness perceptions. Accordingly, H6 is rejected.

5 DISCUSSION

Concerning H1 and H3, simplicity in decision trees for APP positively influenced perceptions of informational fairness mediated by causability. This is consistent with explanatory coherence theory [83] suggesting more simple explanations are preferred, and it is thus crucial to balance providing sufficient information and avoiding overwhelming students with excessive details. Confirming H3, we find evidence for the assumption that more information does not necessarily lead to a better understanding and higher fairness perceptions by default [7, 23], highlighting the importance of students' self-assessed understandability of the decision trees that visualized the ML model's outcome [40]. Moreover, H5, positing institutional trust as a moderator, was not supported, highlighting that trust cannot replace individual comprehension to ensure informational fairness [80]. Regarding RQ1, simplicity had a non-hypothesized positive effect on informational and distributive fairness mediated by causability. This supports existing literature arguing that overall simpler explanations are favored [69, 81, 83, 87].

H2 and H4, which posited that accuracy would impact distributive fairness and be mediated by causability, were not supported. This raises questions about students' awareness of the risks associated with low-accuracy AI. The absence of an effect from a 30-percentage point accuracy manipulation is notable but may also be linked to external validity, as respondents faced no real-world consequences from model errors. Like informational fairness, causability was a strong determinant in distributive fairness, emphasizing its key role in perceptions of fair AI decisions. In the absence of both the overall effects of accuracy on distributive fairness and the indirect effects via causability, H6 had to be rejected. However, institutional trust did affect the relationship between causability and distributive fairness. This suggests that higher institutional trust reduces the importance of the decision tree's understandability on perceptions of distributive fairness. Accordingly, institutional trust may still contribute to APP being judged fairly even if the AI system is not entirely understood, as there can be trust that the university will act ethically and respect students' interests [45, 80]. Lastly, as questioned in RQ2, accuracy also shows no influence on informational fairness. This finding remains constant when adding causability as a mediator in the model, indicating no significant effect. While the concrete level of accuracy of the performance of the APP model does not help to increase informational fairness, this does not mean that the knowledge about the APP's accuracy per se is not important in terms of system transparency. However, since Conijn et al. found that accuracy has no effect on student motivation or confidence in an essay grading system[19], one might assume that accuracy is more critical to the question of whether APP is good enough to be used at all than to the question of whether accuracy increases informational fairness perceptions.

Overall, the results underscore the limited impact of objective attributes like simplicity and accuracy on fairness perceptions, emphasizing the role of subjective factors like causability. Future research should focus on these subjective perceptions to better understand fairness in APP.

6 IMPLICATIONS AND CONCLUSION

6.1 Practical Implications

Our study finds that the design features of the decision tree had a limited impact on its perceived comprehensibility (i.e., causability) and fairness. However, design decisions remain critical for fair and effective communication in XAI. Relevant, high-quality explanations are essential for understanding the APP process and its

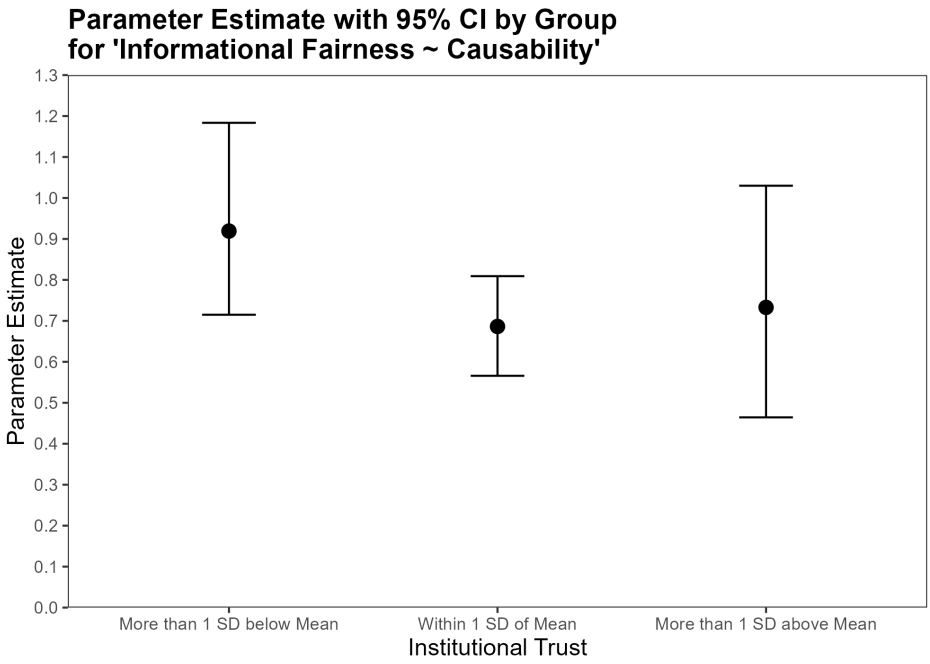

Figure 4: The Effect of Causability on Informational Fairness Depending on Institutional Trust

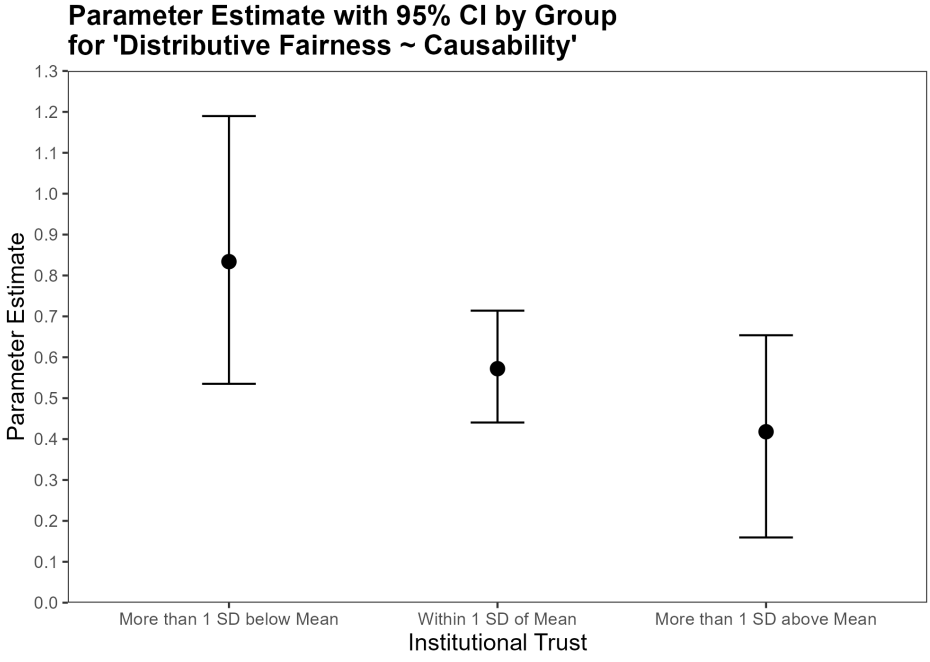

Figure 5: The Effect of Causability on Distributive Fairness Depending on Institutional Trust

outcomes. Yet, these should be presented without overwhelming complexity. Our findings suggest that decision trees with two to five levels do not adversely affect self-assessed understanding or perceived fairness. Overall, respondents indicated that they understood the cause-effect relationships in the white-box model. It would be interesting to explore whether the complexity of other approaches, such as random forest classifiers, could positively or

negatively impact causability. However, it is vital to tailor the complexity of explanations to the audience's cognitive abilities and prior knowledge, especially in diverse educational contexts. Failure to do so risks exacerbating inequalities by favoring those with a greater prior understanding of AI systems. It would thus be premature to dismiss the demands for accuracy regarding these models, as high accuracy is equated with the reliability and trustworthiness of AI systems. Nonetheless, it is essential to acknowledge that well-intentioned policy changes to improve system transparency and fairness may not resonate across all audience segments by default. In sum, the key is providing explanations *and* ensuring they are perceived as comprehensible to those they affect.

6.2 Research Implications

Our study points to the nuanced roles of simplicity and accuracy in shaping fairness perceptions in APP, urging further investigation into other XAI attributes [71]. For example, how counterintuitive factors in decision trees, like high grades predicting poor performance, influence student perceptions remains an open question. The study also calls for understanding how varying student attributes affect the comprehensibility of explanations, raising the issue of whether a one-size-fits-all explanation is adequate. Further research is needed to ascertain the optimal level of explanation that avoids information overload in increasingly complex models. Moreover, the role of universities in APP implementation requires further examination. While institutional trust is vital, it cannot replace the need for individual comprehension of LA explanations. Our findings suggest that decision tree simplicity positively affects fairness perceptions, mediated by causability, whereas prediction accuracy has a less pronounced impact. Our study thus offers critical insights for stakeholders in HEIs, highlighting the importance of balancing explainability and comprehensibility to foster ethical and equitable practices in academic settings.

In conclusion, our study underscores the importance of explainability and comprehensibility in implementing LA and APP technologies in HEIs. Decision tree simplicity emerges as a factor positively influencing fairness perceptions, mediated by causability, while prediction accuracy appears to play a less significant role in shaping student perceptions of fairness. Importantly, our findings highlight the need to explore further institutional trust's multifaceted influence on fairness perceptions in academic contexts. As HEIs strive for fair and transparent APP systems, our research offers valuable insights for policymakers, administrators, and students, fostering ethical and equitable practices in higher education.

ACKNOWLEDGMENTS

Ethics Statement

In conducting this research, we adhered to the highest standards of ethical integrity and responsibility. We ensured that all participants were fully informed about the nature and purpose of the research and provided their informed consent. Participant confidentiality and data privacy were rigorously maintained throughout the study. Ethical guidelines, including those pertaining to non-discrimination, fairness, and respect for individuals, were strictly followed. The research methods were designed to minimize potential harm or discomfort to participants. Any conflicts of interest were disclosed and managed appropriately. This study received approval from the Ethical Review Board of the Faculty of Philosophy of the Heinrich Heine University Düsseldorf, Germany, and was conducted in accordance with international standards for ethical research, such as the Helsinki Declaration.

Funding

This study was conducted as part of the project *Responsible Academic Performance Prediction* (RAPP). The project is funded by the German Federal Ministry of Education and Research [grant number 16DHB4020].

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Availability of Data, Code, and Material

Data and R code for analysis and the Markdown manuscript is provided via OSF, using the same repository where our preregistration details are found: https://osf.io/6td2v/

Declaration of Generative AI in Scientific Writing

During the preparation of this work, the authors used ChatGPT in order to produce and adjust R and Markdown code for the statistical analysis and the reproducible manuscript. We also used ChatGPT to revise and shorten parts of our written draft of the manuscript. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

REFERENCES

[1] Amina Adadi and Mohammed Berrada. 2018. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access* 6 (2018), 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

[2] Deepti Aggarwal, Sonu Mittal, and Vikram Bali. 2021. Significance of Non-Academic Parameters for Predicting Student Performance Using Ensemble Learning Techniques. *International Journal of System Dynamics Applications* 10, 3 (2021), 38–49. https://doi.org/10.4018/IJSDA.2021070103

[3] Khlood Ahmad, Mohamed Abdelrazek, Chetan Arora, Muneera Bano, and John Grundy. 2023. Requirements Practices and Gaps When Engineering Human-Centered Artificial Intelligence Systems. *Applied Soft Computing* 143 (2023), 1–15. https://doi.org/10.1016/j.asoc.2023.110421

[4] Balqis Albreiki, Nazar Zaki, and Hany Alashwal. 2021. A Systematic Literature Review of Student' Performance Prediction Using Machine Learning Techniques. *Education Sciences* 11, 9 (2021), 552. https://doi.org/10.3390/educsci11090552

[5] Sarah Alturki, Ioana Hulpuș, and Heiner Stuckenschmidt. 2022. Predicting Academic Outcomes: A Survey from 2007 Till 2018. *Technology, Knowledge and Learning* 27, 1 (2022), 275–307. https://doi.org/10.1007/s10758-020-09476-0

[6] Eyman Alyahyan and Dilek Düştegör. 2020. Predicting academic success in higher education: literature review and best practices: literature review and best practices. *International Journal of Educational Technology in Higher Education* 17, 1 (2020), 1–21. https://doi.org/10.1186/s41239-020-0177-7

[7] Mike Ananny and Kate Crawford. 2018. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society* 20, 3 (2018), 973–989. https://doi.org/10.1177/1461444816676645

[8] Alessa Angerschmid, Jianlong Zhou, Kevin Theuermann, Fang Chen, and Andreas Holzinger. 2022. Fairness and Explanation in AI-Informed Decision Making. *Machine Learning and Knowledge Extraction* 4, 2 (2022), 556–579. https://doi.org/10.3390/make4020026

[9] Ryan S. Baker and Aaron Hawn. 2022. Algorithmic Bias in Education. *International Journal of Artificial Intelligence in Education* 32, 4 (2022), 1052–1092. https://doi.org/10.1007/s40593-021-00285-9

[10] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, and Francisco Herrera. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion* 58 (2020), 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

[11] Reuben Binns, Max van Kleek, Michael Veale, Ulrik Lyngs, Jun Zhao, and Nigel Shadbolt. 2018. 'It's Reducing a Human Being to a Percentage': Perceptions of

Justice in Algorithmic Decisions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*, Regan Mandryk, Mark Hancock, Mark Perry, and Anna Cox (Eds.). ACM, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3173951

[12] Joana R. Casanova, Cristiano Mauro Assis Gomes, Ana B. Bernardo, José Carlos Núñez, and Leandro S. Almeida. 2021. Dimensionality and reliability of a screening instrument for students at-risk of dropping out from Higher Education. *Studies in Educational Evaluation* 68 (2021), 100957. https://doi.org/10.1016/j.stueduc.2020.100957

[13] Teresa Cerratto Pargman and Cormac McGrath. 2021. Mapping the Ethics of Learning Analytics in Higher Education: A Systematic Literature Review of Empirical Research. *Journal of Learning Analytics* 8, 2 (2021), 123–139. https://doi.org/10.18608/jla.2021.1

[14] Cecilia Ka Yuk Chan. 2023. A Comprehensive AI Policy Education Framework for University Teaching and Learning. *International Journal of Educational Technology in Higher Education* 20, 1 (2023), 1–25. https://doi.org/10.1186/s41239-023-00408-3

[15] Gordon W. Cheung and Roger B. Rensvold. 2002. Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal* 9, 2 (2002), 233–255. https://doi.org/10.1207/S15328007SEM0902{_}5

[16] Thomas K.F. Chiu, Qi Xia, Xinyan Zhou, Ching Sing Chai, and Miaoting Cheng. 2023. Systematic Literature Review on Opportunities, Challenges, and Future Research Recommendations of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence* 4 (2023), 1–15. https://doi.org/10.1016/j.caeai.2022.100118

[17] Jason A. Colquitt and Jessica B. Rodell. 2015. Measuring Justice and Fairness. In *The Oxford Handbook of Justice in the Workplace*, Russell S. Cropanzano and Maureen L. Ambrose (Eds.). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199981410.013.8

[18] Cristina Conati, Kaska Porayska-Pomsta, and Manolis Mavrikis. 2018. AI in Education Needs Interpretable Machine Learning: Lessons from Open Learner Modelling. In *Proceedings of the 2018 ICML Workshop on Human Interpretability in Machine Learning (WHI 2018)*, Been Kim, Kush R. Varshney, and Adrian Weller (Eds.). 267–280. https://doi.org/10.48550/arXiv.1807.00154

[19] Rianne Conijn, Patricia Kahr, and Chris Snijders. 2023. The Effects of Explanations in Automated Essay Scoring Systems on Student Trust and Motivation. *Journal of Learning Analytics* 10, 1 (2023), 37–53. https://doi.org/10.18608/jla.2023.7801

[20] Ricardo Costa-Mendes, Frederico Cruz-Jesus, Tiago Oliveira, and Mauro Castelli. 2021. Machine Learning Bias in Predicting High School Grades: A Knowledge Perspective: A Knowledge Perspective. *Emerging Science Journal* 5, 5 (2021), 576–597. https://doi.org/10.28991/esj-2021-01298

[21] Nelson Cowan. 2010. The Magical Mystery Four: How is Working Memory Capacity Limited, and Why? *Current Directions in Psychological Science* 19, 1 (2010), 51–57. https://doi.org/10.1177/0963721409359277

[22] Russell Cropanzano, Deborah E. Rupp, Carolyn J. Mohler, and Marshall Schminke. 2001. Three roads to organizational justice. In *Research in Personnel and Human Resources Management*, G. R. Ferris (Ed.). Vol. 20. Emerald Group Publishing Limited, Bingley, 1–113. https://doi.org/10.1016/S0742-7301(01)20001-2

[23] Karl de Fine Licht and Jenny de Fine Licht. 2020. Artificial intelligence, transparency, and public decision-making: Why explanations are key when trying to produce perceived legitimacy. *AI & SOCIETY* 35, 4 (2020), 917–926. https://doi.org/10.1007/s00146-020-00960-w

[24] Berkeley J. Dietvorst, Joseph P. Simmons, and Cade Massey. 2015. Algorithm aversion: people erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General* 144, 1 (2015), 114–126. https://doi.org/10.1037/xge0000033

[25] Jonathan Dodge, Q. Vera Liao, Yunfeng Zhang, Rachel K. E. Bellamy, and Casey Dugan. 2019. Explaining Models: An Empirical Study of How Explanations Impact Fairness Judgment. In *Proceedings of the 24th International Conference on Intelligent User Interfaces (IUI '19)*, Wai-Tat Fu, Shimei Pan, Oliver Brdiczka, Polo Chau, and Gaelle Calvary (Eds.). ACM, New York, NY, USA, 275–285. https://doi.org/10.1145/3301275.3302310

[26] Jannik Dunkelau and Manh Khoi Duong. 2022. Towards Equalised Odds as Fairness Metric in Academic Performance Prediction. In *FATED'22: 2nd Workshop on Fairness, Accountability, and Transparency in Educational Data*. 1–6. https://doi.org/10.48550/arXiv.2209.14670

[27] Manh Khoi Duong, Jannik Dunkelau, José Andrés Cordova, and Stefan Conrad. 2023. RAPP: A Responsible Academic Performance Prediction Tool for Decision-Making in Educational Institutes. In *BTW 2023*, Birgitta König-Ries, Stefanie Scherzinger, Wolfgang Lehner, and Gottfried Vossen (Eds.). Köllen, Bonn. https://doi.org/10.18420/BTW2023-29

[28] Chris Edwards. 2021. Let the Algorithm Decide? *Commun. ACM* 64, 6 (2021), 21–22. https://doi.org/10.1145/3460216

[29] Sina Fazelpour and David Danks. 2021. Algorithmic bias: Senses, sources, solutions. *Philosophy Compass* 16, 8 (2021), 1–16. https://doi.org/10.1111/phc3.12760

[30] Heike Felzmann, Eduard Fosch Villaronga, Christoph Lutz, and Aurelia Tamò-Larrieux. 2019. Transparency You Can Trust: Transparency Requirements for Artificial Intelligence Between Legal Norms and Contextual Concerns. *Big Data & Society* 6, 1 (2019), 1–14. https://doi.org/10.1177/2053951719860542

[31] Claes Fornell and David F. Larcker. 1981. Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research* 18, 1 (1981), 39–55. https://doi.org/10.2307/3151312

[32] Egle Gedrimiene, Ismail Celik, Kati Mäkitalo, and Hanni Muukkonen. 2023. Transparency and Trustworthiness in User Intentions to Follow Career Recommendations from a Learning Analytics Tool. *Journal of Learning Analytics* 10, 1 (2023), 54–70. https://doi.org/10.18608/jla.2023.7791

[33] Amit K. Ghosh, Thomas W. Whipple, and Glenn A. Bryan. 2001. Student Trust and its Antecedents in Higher Education. *The Journal of Higher Education* 72, 3 (2001), 322–340. https://doi.org/10.1080/00221546.2001.11777097

[34] Bryce Goodman and Seth Flaxman. 2017. European Union Regulations on Algorithmic Decision Making and a "Right to Explanation". *AI Magazine* 38, 3 (2017), 50–57. https://doi.org/10.1609/aimag.v38i3.2741

[35] David Gunning, Mark Stefik, Jaesik Choi, Timothy Miller, Simone Stumpf, and Guang-Zhong Yang. 2019. XAI - Explainable Artificial Intelligence. *Science Robotics* 4, 37 (2019). https://doi.org/10.1126/scirobotics.aay7120

[36] Alaa Khalaf Hamoud, Ali Salah Hashim, and Wid Aqeel Awadh. 2018. Predicting Student Performance in Higher Education Institutions Using Decision Tree Analysis. *International Journal of Interactive Multimedia and Artificial Intelligence* 5, 2 (2018), 26. https://doi.org/10.9781/ijimai.2018.02.004

[37] Martin Hlosta, Christothea Herodotou, Tina Papathoma, Anna Gillespie, and Per Bergamin. 2022. Predictive Learning Analytics in Online Education: A Deeper Understanding Through Explaining Algorithmic Errors. *Computers and Education: Artificial Intelligence* 3 (2022), 1–12. https://doi.org/10.1016/j.caeai.2022.100108

[38] Kenneth Holstein and Shayan Doroudi. 2022. Equity and Artificial Intelligence in Education. In *The Ethics of Artificial Intelligence in Education*, Wayne Holmes and Kaśka Porayska-Pomsta (Eds.). Routledge, New York, 151–173. https://doi.org/10.4324/9780429329067-9

[39] Andreas Holzinger, André Carrington, and Heimo Müller. 2020. Measuring the Quality of Explanations: The System Causability Scale (SCS): Comparing Human and Machine Explanations: Comparing Human and Machine Explanations. *Künstliche Intelligenz* 34, 2 (2020), 193–198. https://doi.org/10.1007/s13218-020-00636-z

[40] Andreas Holzinger, Georg Langs, Helmut Denk, Kurt Zatloukal, and Heimo Müller. 2019. Causability and Explainability of Artificial Intelligence in Medicine. *Data Mining and Knowledge Discovery* 9, 4 (2019), 1–13. https://doi.org/10.1002/widm.1312

[41] Iddrisu Issah, Obed Appiah, Peter Appiahene, and Fuseini Inusah. 2023. A Systematic Review of the Literature on Machine Learning Application of Determining the Attributes Influencing Academic Performance. *Decision Analytics Journal* 7 (2023), 100204. https://doi.org/10.1016/j.dajour.2023.100204

[42] Weijie Jiang and Zachary A. Pardos. 2021. Towards Equity and Algorithmic Fairness in Student Grade Prediction. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (AIES '21)*, Marion Fourcade, Benjamin Kuipers, Seth Lazar, and Deirdre Mulligan (Eds.). ACM, New York, NY, USA, 608–617. https://doi.org/10.1145/3461702.3462623

[43] Anna Jobin, Marcello Ienca, and Effy Vayena. 2019. The global landscape of AI ethics guidelines. *Nature Machine Intelligence* 1, 9 (2019), 389–399. https://doi.org/10.1038/s42256-019-0088-2

[44] Kyle M. L. Jones, Andrew Asher, Abigail Goben, Michael R. Perry, Dorothea Salo, Kristin A. Briney, and M. Brooke Robertshaw. 2020. "We're being tracked at all times": Student perspectives of their privacy in relation to learning analytics in higher education. *Journal of the Association for Information Science and Technology* 71, 9 (2020), 1044–1059. https://doi.org/10.1002/asi.24358

[45] Kyle M. L. Jones, Alan Rubel, and Ellen LeClere. 2020. A Matter of Trust: Higher Education Institutions as Information Fiduciaries in an Age of Educational Data Mining and Learning Analytics. *Journal of the Association for Information Science and Technology* 71, 10 (2020), 1227–1241. https://doi.org/10.1002/asi.24327

[46] Birte Keller, Marco Lünich, and Frank Marcinkowski. 2022. How Is Socially Responsible Academic Performance Prediction Possible? Insights From a Concept of Perceived AI Fairness. In *Strategy, Policy, Practice, and Governance for AI in Higher Education Institutions*, Fernando Almaraz-Menéndez, Alexander Maz-Machado, Carmen López-Esteban, and Cristina Almaraz-López (Eds.). IGI Global, 126–155. https://doi.org/10.4018/978-1-7998-9247-2.ch006

[47] Hassan Khosravi, Simon Buckingham Shum, Guanliang Chen, Cristina Conati, Yi-Shan Tsai, Judy Kay, Simon Knight, Roberto Martinez-Maldonado, Shazia Sadiq, and Dragan Gašević. 2022. Explainable Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence* 3 (2022), 1–22. https://doi.org/10.1016/j.caeai.2022.100074

[48] René F. Kizilcec. 2016. How Much Information? Effects of Transparency on Trust in an Algorithmic Interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, Jofish Kaye, Allison Druin, Cliff Lampe, Dan Morris, and Juan Pablo Hourcade (Eds.). ACM, New York, NY, USA, 2390–2395. https://doi.org/10.1145/2858036.2858402

[49] René F. Kizilcec and Hansol Lee. 2022. Algorithmic Fairness in Education. In *The Ethics of Artificial Intelligence in Education*, Wayne Holmes and Kaśka

Porayska-Pomsta (Eds.). Routledge, New York, 174–202. https://doi.org/10.4324/9780429329067-10

[50] S. B. Kotsiantis. 2013. Decision Trees: A Recent Overview. *Artificial Intelligence Review* 39, 4 (2013), 261–283. https://doi.org/10.1007/s10462-011-9272-4

[51] Charles Lang and Laura Davis. 2023. Learning Analytics and Stakeholder Inclusion: What do We Mean When We Say "Human-Centered"?. In *LAK23: 13th International Learning Analytics and Knowledge Conference (ACM Digital Library)*, Isabel Hilliger, Hassan Khosravi, Bart Rienties, and Shane Dawson (Eds.). Association for Computing Machinery, New York, NY, United States, 411–417. https://doi.org/10.1145/3576050.3576110

[52] Min Kyung Lee. 2018. Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society* 5, 1 (2018), 205395171875668. https://doi.org/10.1177/2053951718756684

[53] Warren Li, Kaiwen Sun, Florian Schaub, and Christopher Brooks. 2022. Disparities in Students' Propensity to Consent to Learning Analytics. *International Journal of Artificial Intelligence in Education* 32, 3 (2022), 564–608. https://doi.org/10.1007/s40593-021-00254-2

[54] Scott M. Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M. Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. 2020. From Local Explanations to Global Understanding with Explainable AI for Trees. *Nature Machine Intelligence* 2, 1 (2020), 56–67. https://doi.org/10.1038/s42256-019-0138-9

[55] Marco Lünich, Birte Keller, and Frank Marcinkowski. 2023. Fairness of Academic Performance Prediction for the Distribution of Support Measures for Students: Differences in Perceived Fairness of Distributive Justice Norms. *Technology, Knowledge and Learning* (2023), 1–29. https://doi.org/10.1007/s10758-023-09698-y

[56] Marco Lünich and Kimon Kieslich. 2022. Exploring the Roles of Trust and Social Group Preference on the Legitimacy of Algorithmic Decision-Making vs. Human Decision-Making for Allocating COVID-19 Vaccinations. *AI & SOCIETY* (2022), 1–19. https://doi.org/10.1007/s00146-022-01412-3

[57] Hasan Mahmud, A.K.M. Najmul Islam, Syed Ishtiaque Ahmed, and Kari Smolander. 2022. What Influences Algorithmic Decision-Making? A Systematic Literature Review on Algorithm Aversion. *Technological Forecasting and Social Change* 175 (2022), 1–26. https://doi.org/10.1016/j.techfore.2021.121390

[58] Karima Makhlouf, Sami Zhioua, and Catuscia Palamidessi. 2021. Machine Learning Fairness Notions: Bridging the Gap with Real-World Applications. *Information Processing & Management* 58, 5 (2021), 1–32. https://doi.org/10.1016/j.ipm.2021.102642

[59] 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 *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20)*, Mireille Hildebrandt, Carlos Castillo, Elisa Celis, Salvatore Ruggieri, Linnet Taylor, and Gabriela Zanfir-Fortuna (Eds.). ACM, New York, NY, USA, 122–130. https://doi.org/10.1145/3351095.3372867

[60] Barbara Martinez Neda, Yue Zeng, and Sergio Gago-Masague. 2021. Using Machine Learning in Admissions: Reducing Human and Algorithmic Bias in the Selection Process. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education (SIGCSE '21)*, Mark Sherriff, Laurence D. Merkle, Pamela Cutter, Alvaro Monge, and Judithe Sheard (Eds.). ACM, New York, NY, USA, 1323. https://doi.org/10.1145/3408877.3439664

[61] Bahar Memarian and Tenzin Doleck. 2023. Fairness, Accountability, Transparency, and Ethics (FATE) in Artificial Intelligence (AI) and Higher Education: A Systematic Review. *Computers and Education: Artificial Intelligence* 5 (2023), 1–12. https://doi.org/10.1016/j.caeai.2023.100152

[62] Tim Miller. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence* 267 (2019), 1–38. https://doi.org/10.1016/j.artint.2018.07.007

[63] Shira Mitchell, Eric Potash, Solon Barocas, Alexander D'Amour, and Kristian Lum. 2021. Algorithmic Fairness: Choices, Assumptions, and Definitions. *Annual Review of Statistics and Its Application* 8, 1 (2021), 141–163. https://doi.org/10.1146/annurev-statistics-042720-125902

[64] Jovial Niyogisubizo, Lyuchao Liao, Eric Nziyumva, Evariste Murwanashyaka, and Pierre Claver Nshimyumukiza. 2022. Predicting Student's Dropout in University Classes Using Two-Layer Ensemble Machine Learning S Approach: A Novel Stacked Generalization. *Computers and Education: Artificial Intelligence* 3 (2022), 1–12. https://doi.org/10.1016/j.caeai.2022.100066

[65] Andrea Papenmeier, Gwenn Englebienne, and Christin Seifert. 2019. How Model Accuracy and Explanation Fidelity Influence User Trust. http://arxiv.org/pdf/1907.12652v1

[66] Andrea Papenmeier, Dagmar Kern, Gwenn Englebienne, and Christin Seifert. 2022. It's Complicated: The Relationship between User Trust, Model Accuracy and Explanations in AI. *ACM Transactions on Computer-Human Interaction* 29, 4 (2022), 1–33. https://doi.org/10.1145/3495013

[67] Chris Patterson, Emily York, Danielle Maxham, Rudy Molina, and Paul Mabrey. 2023. Applying a Responsible Innovation Framework in Developing an Equitable Early Alert System. *Journal of Learning Analytics* 10, 1 (2023), 24–36. https://doi.org/10.18608/jla.2023.7795

[68] Vanessa Putnam and Cristina Conati. 2019. Exploring the Need for Explainable Artificial Intelligence (XAI) in Intelligent Tutoring Systems (ITS). In *Joint Proceedings of the ACM IUI 2019 Workshops*. 1–7.

[69] Stephen J. Read and Amy Marcus-Newhall. 1993. Explanatory Coherence in Social Explanations: A Parallel Distributed Processing Account. *Journal of Personality and Social Psychology* 65, 3 (1993), 429–447. https://doi.org/10.1037/0022-3514.65.3.429

[70] Irina Rets, Christothea Herodotou, and Anna Gillespie. 2023. Six Practical Recommendations Enabling Ethical Use of Predictive Learning Analytics in Distance Education. *Journal of Learning Analytics* 10, 1 (2023), 149–167. https://doi.org/10.18608/jla.2023.7743

[71] Maria Riveiro and Serge Thill. 2021. "That's (not) the output I expected!" On the role of end user expectations in creating explanations of AI systems. *Artificial Intelligence* 298 (2021), 103507. https://doi.org/10.1016/j.artint.2021.103507

[72] Paul Rodway and Astrid Schepman. 2023. The Impact of Adopting AI Educational Technologies on Projected Course Satisfaction in University Students. *Computers and Education: Artificial Intelligence* 5 (2023), 1–12. https://doi.org/10.1016/j.caeai.2023.100150

[73] Yves Rosseel. 2012. lavaan : An R Package for Structural Equation Modeling. *Journal of Statistical Software* 48, 2 (2012). https://doi.org/10.18637/jss.v048.i02

[74] Cynthia Rudin. 2019. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nature Machine Intelligence* 1, 5 (2019), 206–215. https://doi.org/10.1038/s42256-019-0048-x

[75] Nadine Schlicker, Markus Langer, Sonja K. Ötting, Kevin Baum, Cornelius J. König, and Dieter Wallach. 2021. What to expect from opening up 'black boxes'? Comparing perceptions of justice between human and automated agents. *Computers in Human Behavior* 122 (2021), 1–16. https://doi.org/10.1016/j.chb.2021.106837

[76] Jakob Schoeffer, Niklas Kuehl, and Yvette Machowski. 2022. "There Is Not Enough Information": On the Effects of Explanations on Perceptions of Informational Fairness and Trustworthiness in Automated Decision-Making. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. ACM, New York, NY, USA, 1616–1628. https://doi.org/10.1145/3531146.3533218

[77] Jakob Schoeffer, Yvette Machowski, and Niklas Kuehl. 2021. A Study on Fairness and Trust Perceptions in Automated Decision Making. http://arxiv.org/pdf/2103.04757v1

[78] Donghee Shin. 2021. The Effects of Explainability and Causability on Perception, Trust, and Acceptance: Implications for Explainable AI. *International Journal of Human-Computer Studies* 146 (2021), 1–10. https://doi.org/10.1016/j.ijhcs.2020.102551

[79] Avital Shulner-Tal, Tsvi Kuflik, and Doron Kliger. 2022. Fairness, explainability and in-between: understanding the impact of different explanation methods on non-expert users' perceptions of fairness toward an algorithmic system. *Ethics and Information Technology* 24, 1 (2022), 1–13. https://doi.org/10.1007/s10676-022-09623-4

[80] Sharon Slade, Paul Prinsloo, and Mohammad Khalil. 2019. Learning Analytics at the Intersections of Student Trust, Disclosure and Benefit. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge (LAK '19) (ACM Other conferences)*. ACM, New York, NY, 235–244. https://doi.org/10.1145/3303772.3303796

[81] Megha Srivastava, Hoda Heidari, and Andreas Krause. 2019. Mathematical Notions vs. Human Perception of Fairness: A Descriptive Approach to Fairness for Machine Learning. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '19)*, Ankur Teredesai, Vipin Kumar, Ying Li, Rómer Rosales, Evimaria Terzi, and George Karypis (Eds.). ACM, New York, NY, USA, 2459–2468. https://doi.org/10.1145/3292500.3330664

[82] Christopher Starke, Janine Baleis, Birte Keller, and Frank Marcinkowski. 2022. Fairness perceptions of algorithmic decision-making: A systematic review of the empirical literature. *Big Data & Society* 9, 2 (2022), 1–16. https://doi.org/10.1177/20539517221115189

[83] Paul Thagard. 1989. Explanatory Coherence. *The Behavioral and Brain Sciences* 12, 3 (1989), 435–467. https://doi.org/10.1017/S0140525X00057046

[84] Rahila Umer, Teo Susnjak, Anuradha Mathrani, and Lim Suriadi. 2021. Current stance on predictive analytics in higher education: opportunities, challenges and future directions. *Interactive Learning Environments* (2021), 1–26. https://doi.org/10.1080/10494820.2021.1933542

[85] Warren J. von Eschenbach. 2021. Transparency and the Black Box Problem: Why We Do Not Trust AI. *Philosophy & Technology* 34, 4 (2021), 1607–1622. https://doi.org/10.1007/s13347-021-00477-0

[86] Sebastian Wollny, Daniele Di Mitri, Ioana Jivet, Pedro Muñoz-Merino, Maren Scheffel, Jan Schneider, Yi-Shan Tsai, Alexander Whitelock-Wainwright, Dragan Gašević, and Hendrik Drachsler. 2023. Students' expectations of Learning Analytics across Europe. *Journal of Computer Assisted Learning* (2023). https://doi.org/10.1111/jcal.12802

[87] Claire Woodcock, Brent Mittelstadt, Dan Busbridge, and Grant Blank. 2021. The Impact of Explanations on Layperson Trust in Artificial Intelligence-Driven Symptom Checker Apps: Experimental Study. *Journal of medical Internet research* 23, 11 (2021), e29386. https://doi.org/10.2196/29386

[88] Kun Yu, Shlomo Berkovsky, Ronnie Taib, Dan Conway, Jianlong Zhou, and Fang Chen. 2017. User Trust Dynamics: An Investitgation Driven by Differences in System Performance. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces*, George A. Papadopoulos, Tsvi Kuflik, Fang Chen, Carlos Duarte, and Wai-Tat Fu (Eds.). ACM, New York, NY, USA, 307–317. https://doi.org/10.1145/3025171.3025219

[89] Mireia Yurrita, Tim Draws, Agathe Balayn, Dave Murray-Rust, Nava Tintarev, and Alessandro Bozzon. 2023. Disentangling Fairness Perceptions in Algorithmic Decision-Making: the Effects of Explanations, Human Oversight, and Contestability. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, Albrecht Schmidt, Kaisa Väänänen, Tesh Goyal, Per Ola Kristensson, Anicia Peters, Stefanie Mueller, Julie R. Williamson, and Max L. Wilson (Eds.). ACM, New York, NY, USA, 1–21. https://doi.org/10.1145/3544548.3581161

[90] Baobao Zhang and Allan Dafoe. 2019. Artificial Intelligence: American Attitudes and Trends. https://doi.org/10.2139/ssrn.3312874

[91] Jianlong Zhou, Fang Chen, and Andreas Holzinger. 2022. Towards Explainability for AI Fairness. In *xxAI - Beyond Explainable AI*, Andreas Holzinger, Randy Goebel, Ruth Fong, Taesup Moon, Klaus-Robert Müller, and Wojciech Samek (Eds.). Springer International Publishing, Cham, 375–386. https://doi.org/10.1007/978-3-031-04083-2{_}18