Fairness and the Need for Regulation of AI in Medicine, Teaching, and Recruiting

Laila Wegner, Yana Houben, Martina Ziefle, and André Calero Valdez(✉)

Human-Computer Interaction Center, RWTH Aachen University, Campus Boulevard 57, 52076 Aachen, Germany
laila.wegner@rwth-aachen.de
yana.houben@kpnplanet.nl
{ziefle,calero-valdez}@comm.rwth-aachen.de

Abstract. The connections between the perception of trust, fairness, and regulatory needs regarding artificial intelligence have not been sufficiently investigated, yet. We address this research gap and analyze the role of trust, acceptance, and confidence in technology use towards the need for regulations and perceptions of fairness of artificial intelligence. A quantitative questionnaire (n = 103) was used to empirically and deductively study the aforementioned research question and established hypotheses. Overall, the results suggest most importantly trust has an impact on assessing the fairness of AI, and that it correlates with regulatory needs. Furthermore, we found that trust and a lack of regulatory needs predict the assessment of perceived AI fairness, explaining 41% of the variance. We also found that the usage context has a significant impact on perceived fairness and regulatory needs. Interestingly, teaching showed the highest need for regulation of all our contexts and showed the lowest perceived fairness.

Keywords: Fairness · Regulatory needs · Artificial intelligence · User study · AI in teaching · AI in medicine · AI in recruiting · Trust in AI · Contextual AI · Technology acceptance

1 Introduction

Artificial intelligence is one of the major topics of our time. Scientific discussions occur in a variety of disciplines, from the perspective of computer scientists to technical feasibility to philosophical and psychological elaborations of social consequences. The prominence of artificial intelligence is also growing in public discourse, and many headlines circle around the topic. For example, "ZEITonline," a German news magazine, reports: "Congratulations, you have convinced the AI! - When it is a matter of who is hired or promoted, prejudices also influence us. Should difficult decisions rather be made by a computer? [...] **Fairness through algorithms—that sounds too good to be true.** Unfortunately, it is. It has become common knowledge that algorithms can also discriminate.

V. G. Duffy (Ed.): HCII 2021, LNCS 12778, pp. 277–295, 2021.
https://doi.org/10.1007/978-3-030-77820-0_21

One case that has gained unfortunate notoriety comes from the field of criminal justice: [...]" [10]

The social context that artificial intelligence can change has long been recognized. Initial use is already being tested in some fields of application. Because the technology is in the development phase, it currently carries some risks. As the quote from DIE ZEIT makes clear, **fairness in the use of artificial intelligence is one of the central aspects in the public discourse**. The above quote has a very negative connotation and provides an example of the fears when it comes to evaluation by algorithms.

The topic of fairness is also already being considered in science. In the following chapter, we look at the state of research, which types of artificial intelligence exist, and how the aspects of fairness are evaluated. As O. Renn points out in his essay, the **establishment of new technologies is closely linked to their societal acceptance** [18].

However, the correlations between the social perception of the fairness of artificial intelligence, technology acceptance, and the establishment of artificial intelligence have not yet been sufficiently investigated. The goal of this elaboration is to be able to make statements about these interrelationships, and to this purpose, the following question is examined: ***What is the effect of technology acceptance and trust in technology on the need for regulation to ensure the fair use of artificial intelligence?***

We derive five hypotheses to answer this research question. To be able to test the hypotheses, we conduct an empirical and quantitative survey. The survey works with the evaluation of different application scenarios. More details regarding the method follow in chapter four. To be able to place the survey responses in a broader context, the sample description follows in the fifth chapter. In the results chapter, the relationships established in the hypotheses are analyzed using the statistical programming language R.

To analyze the results in the context of the research proposal, they are contrasted with the expectations derived from the state of research in the following sections. The research question is answered with the help of our findings.

Out of five hypotheses, three could be accepted. We could confirm that strong trust in AI leads to low regulatory needs. Furthermore, a significant impact of trust and distrust in technology on the need for regulation was shown. The need for regulation for three scenarios was compared and it could be confirmed that they differ depending on each scenario. However, we could not confirm that there is a correlation between technology affinity and the need for regulation in AI. We also were not detecting any effect of injustice sensitivity on the expected fairness of AI. The meaning of these findings as well as the importance for the fair use of artificial intelligence is presented in the conclusion.

2 Related Work

As mentioned at the beginning, artificial intelligence is a technology with interdisciplinary relevance. The research area that has emerged is correspondingly large. To be able to classify the current state, some basics and development trends of artificial intelligence are considered in the following. The main focus here is on the different types of artificial intelligence and its proliferation. Subsequently, this chapter is dedicated to the previous discussions of ethical challenges to be able to delimit which scientific findings have been gained on the fair use of artificial intelligence. To accomplish this, some specific use cases are also discussed thereafter. Finally, some findings from acceptance research and the influence of certain personality traits are considered.

2.1 Trends in Artificial Intelligence Development

A clear definition of the term artificial intelligence (AI) is sometimes difficult to delineate. According to McCharty in 1955 artificial intelligence are machines that behave as if they have human intelligence [12]. However, because human intelligence is also difficult to define, this definition is appropriately abstract. More recent definitions instead work with degrees of intelligence. Mainzer, for example, provides the following working definition in his book "Künstliche Intelligenz – Wann übernehmen die Maschinen"[1]: "A system is called intelligent if it can solve problems independently and efficiently. The degree of intelligence depends on the degree of self-reliance, the degree of complexity of the problem, and the degree of efficiency of the problem-solving procedure (translated from German)." [13]

These different degrees are also reflected in the distinction between strong and narrow (or weak) AI. Narrow AI is an expert in one single area; abstraction into other contexts is not possible. Whereas in the development of strong AI, the goal is for the AI to acquire the same intellectual capabilities of a human. With the current state of the art, all existing systems belong to the category of narrow AI [20]. Furthermore, emotions or empathy cannot be reproduced by an AI, only simulated. However, it is possible to program ethical behavior based on rules and machine learning [9]. One trend that can currently be observed is the spread of ambient intelligence. This describes the networking of sensors, radio modules, and computer processors that are integrated into everyday life and serve to improve it [6].

2.2 Addressing Ethical Challenges

Artificial intelligence is thus gradually spreading into all areas of society. This includes high-risk areas (such as medicine), increasing the relevance that AI must be designed to be fair and transparent. Unforeseen and collateral cultural impacts cannot be ruled out [5]. The development of AI creates new opportunities for the economy. On the one hand, new products and services are conceivable, as they provide an enormous increase in productivity. On the other hand, though, the increased use of AI can also lead to increased unemployment and greater wealth disparities than before [14]. Many commentators, academics, and

[1] Artificial Intelligence - When machines take over.

policymakers are therefore calling for ensuring that algorithms are transparent, fair, and accountable [16]. One possible solution comes from Iyad Rahwan. He demands the regulation of AI and proposes the "programming of an algorithmic social contract". Here, the characteristics of successful algorithmic regulation are based on O'Reilly. These require a deep understanding of the desired outcome; real-time measurement to determine whether that outcome is being achieved; algorithms (i.e., a set of rules) that make adjustments based on new data; and periodic, a deeper analysis of whether the algorithms themselves are correct and working as expected.

Recent policy decisions such as the GDPR[2] in Europe show that the need for action is recognized. These laws provide the initial legal basis to address impacts from AI on society. The focus is on the fair processing of personal data [2].

2.3 Current Use Cases

As mentioned, artificial intelligence is spreading into many different areas. Some use cases are being tested and analyzed. The following briefly outlines three significant use cases, which will also be addressed in the empirical survey.

– **Medicine**: In medicine, AI has the potential to optimize the care pathway for chronically ill patients. Artificial intelligence can be used to plan precise therapies for complex diseases, reduce medical errors, and improve enrollment of subjects in clinical trials. Although absolute confidence in the diagnostic performance of artificial intelligence has not yet been established, the combination of machines and physicians reliably improves system performance [15].
– **Human resource management**: If a new employee is sought, the support of artificial intelligence is possible. This supports the decision through prepared analyses of video interviews. However, in addition to supporting the goal of finding the optimal employee, the use of AI in the HR management process also brings the potential for discrimination. Moreover, potential legal and ethical consequences must be considered [7].
– **Teaching**: In teaching, the increasingly widespread use of eLearning portals can be observed. Acceptance and success of this medium can only be achieved if the systems act as helpful assistants and are not designed to be too complex. Intelligent guidance and situational support for the students are necessary for this. In addition, adaptivity for individual use of the portal is an essential feature [11].

2.4 Acceptance Research

As mentioned at the outset, the spread of new technologies is closely linked to their acceptance. Artificial intelligence is a relatively new phenomenon on which there is still little comprehensive research. However, acceptance research has been conducted in numerous disciplines and some findings can be applied

[2] General Data Protection Regulation.

to the subject area of this study. The strong influence of acceptance on the diffusion of new products lies on the one hand in the fact that the absence of resistance follows from acceptance, and on the other hand in the fact that acceptance leads to active participation and willingness to act. Therefore, **acceptance research also results in approaches for successful technology implementation** [22]. In addition to the usefulness of new products, product acceptance, as well as ethical and moral attitudes and widespread thoughts and beliefs about humanity, also play a central role in the perception of consumers today. Products such as artificial intelligence, which lead to strong individual and social change, can only be realized with broad acceptance [22].

If acceptance is to be promoted, innovation faces the challenge that acceptance is a subjective variable and cannot be enforced. However, it is possible to contribute to an increase in acceptance by tailoring technology to the respective target group and implementing it competently. The measures to achieve this should be aimed at reducing the perceived costs of the technical innovation and increasing the benefits. Recommendations from the literature are based in part directly on the object of acceptance and focus on an adapted design of the new technology [22]. This approach is pursued in the context of this elaboration. By asking potential users about their regulatory needs, it is possible to use the insights gained to adapt the regulations of artificial intelligence and thus increase acceptance.

Trust in technology is also an influential factor. If a potential consumer distrusts a new product, their perception focuses on the risks. With trust, on the other hand, the consumer relies on the satisfaction of their expectations [8].

2.5 Personality Characteristics

As the previous section made clear, acceptance is a subjective factor. Thus, individual personalities also influence the perception of new techniques. Since the focus of this paper is the fairness of artificial intelligence, we analyze the participants' sensitivity to unfairness. It is known from research that different perceptions and reactions to unfairness can be identified. These differences can be generalized across different unfair situations [21].

3 Research Question and Hypotheses

To answer the overarching question in this article we propose the following research question: *RQ: What is the effect of technology acceptance and trust in technology on the need for regulation and how does it impact the perception of fairness of artificial intelligence?* To answer this research question (see Fig. 1), five hypotheses were derived from the state of the art research, which will be answered throughout the paper using the survey. In the following, the established hypotheses are stated and the justified expectations are outlined.

H1: People Who Generally Show Low Technology Affinity Exhibit a High Need for Regulation of AI. This hypothesis focuses on the relationship between technology

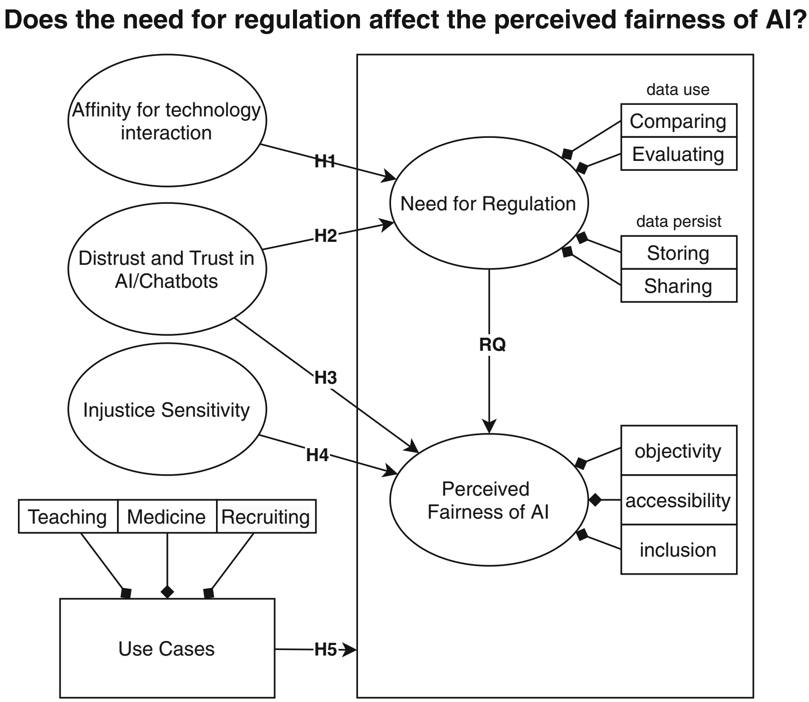

Fig. 1. The proposed research model of this article including hypotheses and measurement model.

acceptance and the need for regulation. From acceptance research it became clear that only accepted products are used, if this acceptance is missing this leads to restrictions in the willingness to use. Therefore, the expectation is: those who are generally not open to new technologies are also averse to artificial intelligence. As a consequence, restrictions on the unaccepted product follow through regulation.

H2: People Who Exhibit Strong Trust in AI Have Low Regulatory Needs When Using AI. The second hypothesis analyzes the factors of technology trust and the need for regulation. Acceptance research indicates that distrust focuses on risks, whereas trust relies on the occurrence of expectations. To manage the risks, regulations would be a possible solution. If trust dominates, this solution is not needed because the risks are not the focus. The expectation is therefore that there is a negative correlation between trust in artificial intelligence and the need for regulation.

H3: Trust and Distrust in Technology Play a Significant Role in the Perceived Fairness of AI. The third hypothesis assumes the perception of fairness of AI can not yet have been established from real interactions, or if so only to a small degree. Therefore, we hypothesize that the attitude towards AI, in this case, trust or distrust, plays a large role in how the fairness of AI is evaluated.

H4: People Who are Highly Sensitive to Justice are Critical of the Fairness of Artificial Intelligence. The fourth hypothesis addresses the fairness aspect of artificial intelligence. As was shown from the research on unfairness sensitivity, differences in sensitivity can be generalized across different situations. The resulting expectation is that this generalization can also be applied to new technologies (such as artificial intelligence).

H5: The Required Need for Regulation Varies Between the Different Use Cases. The research question considers the use of AI. Different application scenarios have already been studied in the literature, and different benefits and risks have been identified in different application scenarios. However, a comparison of user acceptance across application scenarios has not been extensively researched. Therefore, it is of interest whether findings on user acceptance from individual application areas can be abstracted to other areas. Since the benefits and risks have different consequences per use case, the expectation is that the need for regulation will vary depending on the scenario.

4 Method

After the literature review, the research question and five different hypotheses were formulated. With the help of these hypotheses, it should be possible to answer the research question at the end of this quantitative study. Empirical data collection was conducted with the help of a questionnaire to afterward deductively investigate the previously formulated research question and established hypotheses.

Next, we present the designed survey questionnaire, the used scales, as well as the statistical procedures in this study.

4.1 Materials and Survey Design

The questionnaire consists of some demographic questions and three scenarios, with questions on fairness and regulatory needs for each. The order of the scenarios is randomly assigned and all items that should not be sorted were additionally randomly presented to avoid effect errors. The questionnaire contains an introduction as well as a message of gratitude at the end. Participants were informed that data is gathered anonymously and voluntarily.

Before the data were collected, the questionnaire was administered in a pretest with three participants. The time required was recorded. After the questionnaire was improved, data collection was started. A within-subject design was used as the experimental design, which means that each subject had to answer questions on all scenarios, albeit in randomized order. Participants were acquired using a mix of methods between self-selection, snowball effect, and a deliberate selection process. On June 5th 2019, participants were personally contacted by their circle of acquaintances via social media. On June 18th 2019, the questionnaire was closed.

4.2 Description of Measurement Instrument and Scenarios

In the questionnaire, which can also be viewed in the OSF repository, three scenarios are compared. Introductory texts are presented here, translated into English. Original texts are available in the OSF repository as well.

Medicine Scenario. The first scenario, Medicine, is defined as follows:

Imagine an artificial intelligence that you can contact with medical questions. Here, natural written or verbal communication (via keyboard or telephone) with the artificial intelligence is possible. It can answer questions about health and make diagnoses based on the chat or conversation with the patient. For the latter, it is able to ask specific queries. In addition, appointments for a possible subsequent doctor's visit are coordinated by the AI.

Teaching Scenario. The second scenario teaching is introduced with the following text:

Imagine you are a student at a language school. An artificial intelligence that understands natural languages accompanies the lessons. It can respond to the individual knowledge levels of the learners and answer follow-up questions from the students. In addition, it can act as a training partner and apply various pedagogical concepts.

Recruiting Scenario. The third scenario is human resource management, defined by the following description:

Imagine HR management processes at work being supported by an AI. The AI analyzes and evaluates your application documents. In addition, the AI summarizes all the data for the HR manager. In the end, the AI coordinates the interviews, and a chatbot is used to communicate with the applicants.

Regulatory Needs and Fairness. After each scenario, the regulatory needs, adapted to the scenario, are asked by using six-point Likert scales. Fairness was measured on a six-point bipolar scales with textual opposing anchors for levels one and six. Participants are asked to indicate their personal opinion for each regulatory need and for each fairness item. In total, there are five different regulatory needs and three different fairness items per scenario.

The first **regulatory requirement** for each scenario describes data storage. The second regulatory requirement focuses on the comparison of personal data with other users and the third focuses on the analysis of the data. The last regulatory requirement deals with the transfer of data.

The different **fairness** items ask whether users expect to be treated equally or whether they believe that certain groups are expected to be disadvantaged. Next, we ask whether the AI is expected to function safely or whether it can be manipulated and is faulty. The third queried expectation is whether disclosure of information will lead to improvements or to disadvantages for the user. The fourth contrasts adequate evaluation competence by AI against lack of inclusion of individual evaluation aspects, and the last question compares whether benefits by AI are accessible to all users versus benefits are not accessible to all user groups.

4.3 Data Analysis and Statistical Procedure

After the survey, the analysis of the data took place. The data set was reduced to rows of data with at least 50% of the data present. After the incomplete cases were taken out, all items were renamed to prepare for further analysis. All statistical procedures were conducted R Version 4.0.2. All data manipulations were conducted using the `tidyverse` [24]. All procedures are available on a GitHub Repository[3]. Data and supplementary materials are available at an OSF repository[4]. Supplementary materials are created using several R-packages [1,25].

We first verify the internal consistency of existing scales using Cronbach's α. For item sets designed for this study, we use exploratory factor analysis to determine the internal structure of these item sets. Both methods were taken from the R `psych` package [19]. We verify assumptions to factor analysis using Bartlett's test of sphericity and the Kaiser-Meyer-Olkin criterion of sampling adequacy.

Factor calculation was done using the `hcictools` package [4]. Descriptive analysis of relevant variables was conducted using the `psych` package [19].

We tried testing our proposed model using a partial least squares structural equation model from the `seminr` package [17]. However, measures of reliability were not sufficient for our data. The efforts to model our data using `seminr` are available in the supplementary materials.

Next, we conducted correlation analysis with Pearson moment correlation for the variables in question using tools from the `hcictools` package. We used repeated-measures ANOVA and multiple linear regression from the `jmv` package [23].

In general, we assume a level of significance of $\alpha = .05$, meaning that when findings are significant, there is a 5% change that our data could have been observed given the null-hypothesis is true. We use non-parametric tests when we have reason to assume that underlying population data would not be normally distributed.

5 Results

Using the aforementioned statistical methods, we now describe our findings in three sections. First, we describe the data set using descriptive statistics. Next, we test our hypotheses using correlation analysis and repeated-measures ANOVA. Lastly, we use multiple linear regression to determine the impact of our variables on our target variable perceived fairness of AI.

[3] https://github.com/Sumidu/AIFairnessPaperHCII2021.
[4] https://osf.io/54fjy/.

5.1 Sample Description

In any study, both different and similar characteristics of the subjects are of concern. It is useful to list the samples in order to be able to compare the data with the samples afterward.

In total, there were 136 participants in the survey, of which 103 completed more than half of the questionnaire. The mean age is rather young with M=33. The youngest participant is 14 years old and the oldest participant is 77. Out of 103 participants, 58 are women and 45 are men.

School-leaving qualifications were asked and it is shown that most of the respondents have completed at least the vocational diploma (German: Fachabitur/Abitur), so the educational level of the respondents is comparatively high. One person has no school-leaving qualification, one mentions the Certificate of Secondary Education (German: Hauptschulabschluss), two mentions the General Certificate of Secondary Education (German: Realschulabschluss, there were 48 answers with vocational diploma, 15 times vocational training and 36 times university degree.

In Table 1 an overview with the mean of our main variables is shown. With M=3.72, the respondents show a rather high level of trust in AI. Furthermore, with M=3.77, the need for regulation in the use of data is also rather high. It is striking that the need for regulation in data persistence is noticeably lower with M = 2.81.

Table 1. Descriptive overview of our main variables

Variable	n	mean	sd	se
Age	103	33.16	15.43	1.52
Affinity towards technology	103	3.04	1.11	0.11
Injustice sensitivity	103	3.25	0.93	0.09
Need for regulation - data use	103	3.77	1.08	0.11
Need for regulation - data persist	103	2.81	0.98	0.10
Trust in AI	103	3.72	0.77	0.08
Distrust in AI	103	3.57	0.77	0.08

5.2 Hypotheses Tests

In the following section, we will test our hypothesized associations in our model by applying correlation analysis. We assume normality on all scales with more than three items and thus use Pearson's moment correlation for analysis. For our first four hypotheses, we test the effect of our independent variables and the respective dependent variable for all scenarios in a single measure as the single measure achieved the highest reliability and factor analysis often did not yield strong enough variation to assume multiple factors.

H1: People Who Generally Show Low Technology Affinity Exhibit a High Need for Regulation of AI. The need for regulation was measured on two scales. We see no significant correlation of affinity towards technology (ATI) with a need for regulation regarding the use of data ($r(101) = .06, p > .05$). This means that the need for regulation when it comes to using data in all our scenarios does not depend on the individual's affinity towards technology. The same is true for a need for regulation regarding the persistence of data ($r(101) = -.15, p > .05$).

H2: People Who Exhibit Strong Trust in AI Have Low Regulatory Needs When Using AI. Our trust-related items showed an interesting two-factor structure meaning that trust in AI and distrust in AI are not complete opposites on the same scale. Exploratory factor analysis yielded two factors that are negatively correlated on their primary axes ($r = -.424$). The scores themselves, as expected, also show a negative correlation ($r = -.33$)

With this, it is interesting to see that trust plays a larger role in the need for regulation than distrust in our sample (see Fig. 2). We see no correlation between distrust and both measures for the need for regulation ($|r| < .13, p > .05$). Meaning that a general distrust in AI does not translate to a stronger need for regulation directly. However, trust is correlated with both measures. It is weakly correlated with the need for regulation regarding data persistence ($r(101) = -.22, p < .05$), meaning that the higher the users trust in AI the less they are concerned about data storage. This effect is even larger for data usage ($r(101) = -.31, p < .01$). Here, a medium effect is seen, meaning that the more a user trusts AI in general the less they are worried about the use of data by AI.

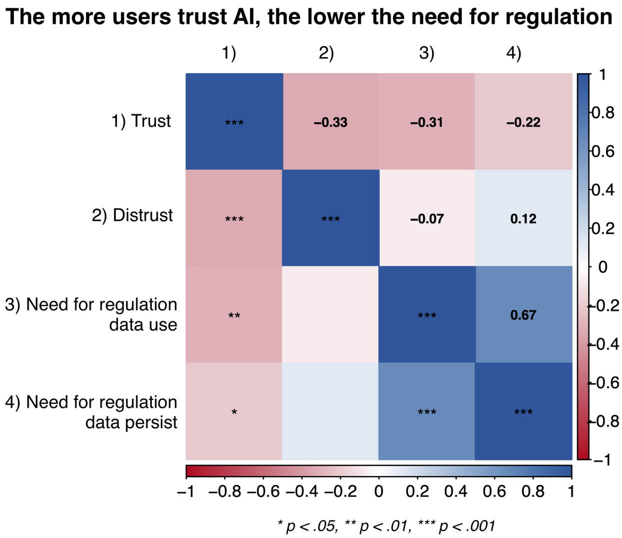

Fig. 2. Evaluating how users perceptions of AI related to the need for regulation.

H3: Trust and Distrust in a Technology Play a Significant Role in the Perceived Fairness of AI. As trust towards a trusted subject is associated with fair behavior we also test whether trust in AI plays a role in the perceived fairness of AI. Indeed we find that trust is positively associated with perceived fairness ($r(101) = .41, p < .001$). This means that users that show high trust in technology also expect AI to be fairer than users that show lower trust in technology. Interestingly, the inverse measure—distrust—is also correlated with fairness, yet at a lower effect ($r(101) = -.26, p < .01$).

H4: People Who Are Highly Sensitive to Justice Are Critical of the Fairness of Artificial Intelligence. One of our assumptions was that the perception of fairness of AI would also be connected to a general sensitivity towards injustice. If a person is more aware of injustice by being more sensitive, they should also be more attuned to detecting injustice and thus the lack of fairness in an AI system. However, this association is only very weak ($r(102) = -.16, p > .05$) and not statistically significant. We must assume that participants that are more sensitive to injustice do not expect AI to be more unfair—in our scenarios at least.

H5: The Required Need for Regulation Varies Between the Different Use Cases. Although factor analysis did not reveal a clear factor structure in our dependent variables from all three scenarios, we still can investigate differences in means between the different scenarios. The resulting short scales also show high reliability, but they are highly correlated. A shift in average evaluation between scenarios is still imaginable and thus tested here.

We use a repeated-measures ANOVA to test for differences between scenario choices and both fairness and need for regulation (both scales combined).

For fairness, we see that the ANOVA yields significant differences between contexts ($F(2, 204) = 6.75, p < .001$). Post-hoc Tukey corrected tests show that there is only a difference in means between the teaching and the recruiting scenario ($t(204) = -3.64, p < .001$, see supplementary materials for the full ANOVA tables). The expected fairness is highest in recruiting ($M = 3.52, SE = 0.104$), and lowest in teaching ($M = 3.14, SE = 0.104$). The expected fairness is on a medium level for the medicine scenario ($M = 3.38, SE = 0.104$). It is important to note that the overall fairness is rather low (below or near the scale mid-point of 3.5). For visual inspection, we plotted the means and the Cosineau-Morey within-subject confidence intervals (95% CIs) in Fig. 3.

For the need for regulation (both scales combined) we also significant differences between usage scenarios ($F(2, 204) = 37.4, p < .001$). Here, all scenarios show different means when looking at Tukey-corrected p-values (all $p < .05$). Interestingly, the need for regulation in medicine scores lowest ($M = 2.92, SE = 0.09$), while the need for regulation in teaching scores highest ($M = 3.57, SE = 0.09$). Recruiting is in the middle place with a mean of 3.16 ($SE = 0.09$).

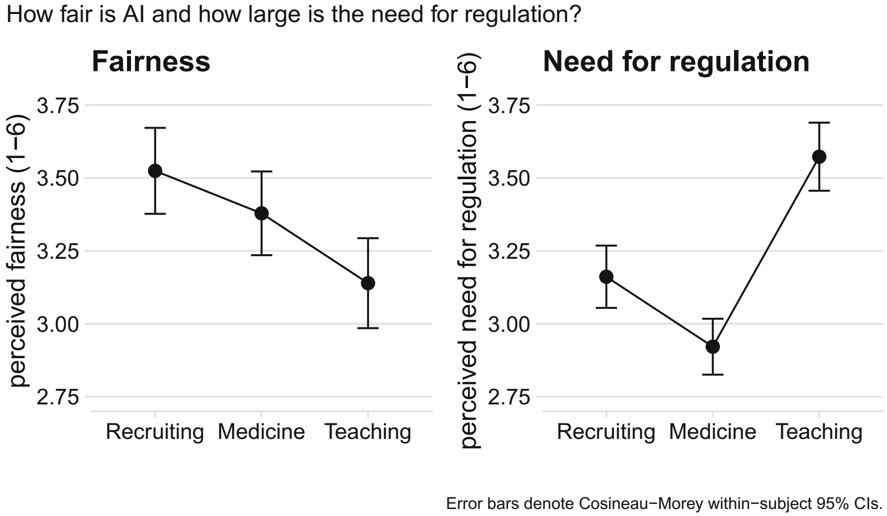

Fig. 3. Comparing perceived fairness and the need for regulation between contexts.

When viewed together with the fairness findings it is interesting to see that teaching plays a different role in the evaluation of fairness and the need for regulation.

5.3 Main Research Question

The main research question of this article was whether the need for regulation impacts the perceived fairness of AI in different contexts. We have already seen that the teaching context might be peculiar in our data, and therefore resort to investigate this effect using the combined factor scales for fairness and the two scales for the need for regulation. We use multiple linear regression to test whether the other correlated variables (trust, injustice sensitivity) also impact the perceived fairness. We do this by using the enter method and comparing three models.

The first mode uses both regulatory needs as predictors. The second model adds trust as a predictor and the third adds injustice sensitivity. The multiple regression analysis showed that the initial two-predictor approach was sufficient in explaining 37% of the variance in the perceived fairness ($F(2, 100) = 29.6, p <$ .001). Adding trust into the equation increases the explained variance only to 43% ($F(3, 99) = 25.7$), but at the same time causing the 0 to inside the 95% confidence interval for the coefficient of data use regulatory needs. Adding injustice sensitivity did not make two predictors become not significant.

Therefore, we also tested a model that uses data persistence and trust to predict fairness. This model able to explain 41% of the variance ($F(2, 100) = 36, p < .001$, see Table 2). The need for regulation had a standardized coefficient of –0.51, while trust had a standardized coefficient of 0.3.

Overall, we can say that the perceived fairness of AI is thus strongly influenced by the need for regulation regarding data persistence and weakly influenced by a general trust in AI.

Table 2. Linear regression table for perceived fairness of AI

Variable	B	SE	Stat	t	p
(Intercept)	3.66	0.42	8.80	8.80	<.001
Need for regulation - data persist	−0.45	0.07	−6.54	−6.54	<.001
Trust in AI	0.33	0.09	3.84	3.84	<.001

We can also look at the individual fairness ratings for each context (see Fig. 4). Here, we look at whether to mean rating of a value is significantly different from the scale mean of 3.5. We can see that our participants do believe that AI is able to assess personal skills more objectively, while at the same time thinking that the evaluation of soft skills was maybe insufficient. In the medical scenario, users are afraid that data is used to their disadvantage from insurance companies and that AI might make incorrect diagnoses from data. However, they do think AI can help in making better and fairer appointment schedules for patients. In the teaching scenario, participants do believe that the quality of teaching will improve fairly and that an AI system will be able to evaluate learning progress more objectively.

6 Discussion

After presenting our findings, we contextualize our results in light of other research. We first look at the individual hypotheses before discussing the implications of the findings regarding the main research question.

6.1 Discussion of the Hypotheses

The first hypothesis *People who generally show low technology affinity exhibit a high need for regulation of AI* was not found to be true and could therefore be rejected. According to the state of research, only accepted products are used. If this acceptance is missing, this leads to restrictions in the willingness to use. This leads to the expectation that people who are generally not open to new technologies will also be negative towards artificial intelligence and that this will result in restrictions on the unaccepted product through regulation. Since the hypothesis could be rejected, the result does not match the expectation derived from the current state of research. One possible reason that technology acceptance has no influence on regulatory needs is that the respondents assess technology acceptance based on current technical products and therefore cannot directly imagine artificial intelligence as a technical product and there is, therefore, no correlation

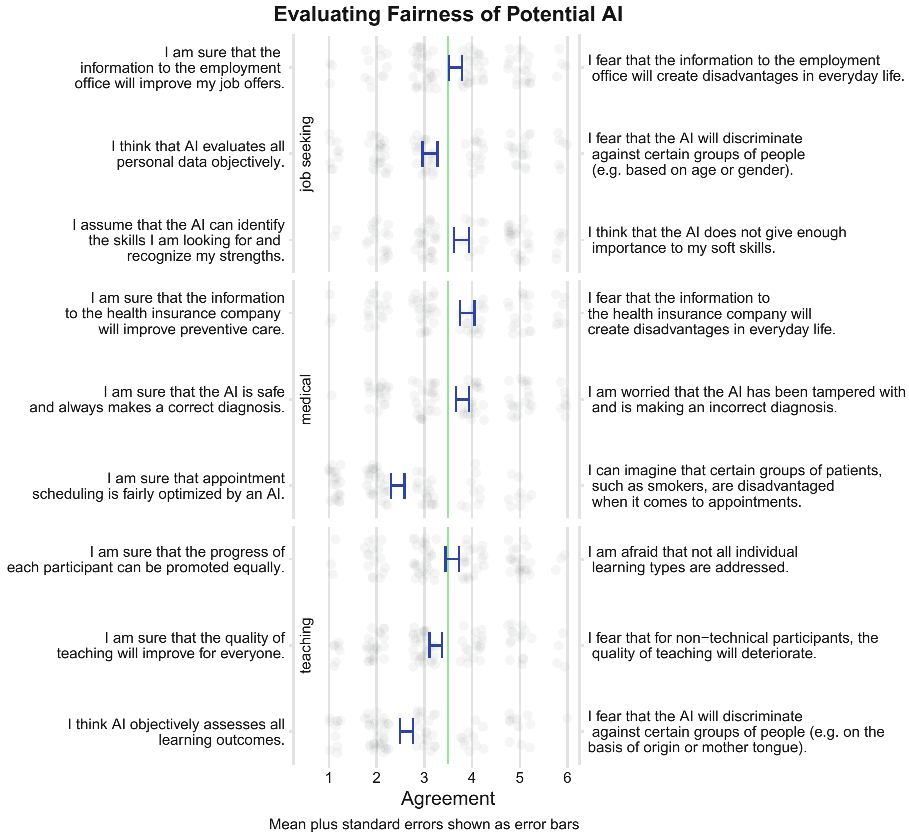

Fig. 4. Comparison of fairness evaluations for all contexts

between the two variables. To confirm or re-examine this, another study should be conducted in the future that examines the acceptance of artificial intelligence and not only the general acceptance of technology.

According to the state of research, for people with low trust in technology, regulating AI is a possible solution to improve their trust in AI. As an expectation, it follows that there is a negative correlation between trust in AI and desired regulatory needs, as well as fairness.

The second and third hypotheses *People who exhibit strong trust in AI have low regulatory needs when using AI and perceive AI as more fair* were found to be true. Thus, the result is consistent with the expectation.

Interestingly, we found that trust and distrust towards AI are not full antonyms using factor analysis. This indicates the multi-faceted nature of AI and that future evaluation should focus on more detailed aspects of AI when evaluating trust and distrust. Moreover, it is recommended to study the general propensity to trust both individuals and technology as possible confounds in future research.

Unexpectedly, we found that the persistence of data played a larger role in determining the fairness of AI than the use of data. Other studies in the field of privacy research have found conflicting evidence for this finding [3]. Here, secondary use is considered particularly harmful for acceptance. However, we did not ask for "secondary" use, but for primary use, which could explain the increased importance of data persistence in our scenarios.

Another expectation based on the current state of research was that injustice sensitivity can be generalized to different situations and that this generalization can also be applied to new technologies such as artificial intelligence. The fourth hypothesis *People who are highly sensitive to justice are critical of the fairness of artificial intelligence* followed from this expectation and could not be found to be true and therefore is rejected. This means that the expectation is not consistent with the hypothesis. One possible reason for this would be that respondents do not (yet) associate injustice sensitivity with technical situations. For example, people could not be able to imagine different ways that AI can treat users unfairly. This lack of "negative creativity" could have caused users with higher injustice sensitivity to react similarly to users with lower injustice sensitivity. At the same time, it would be possible that people find it difficult to imagine AI and thus evaluate the trust in AI, the fairness of AI, and the desired need for regulation with the same tendency every time.

Furthermore, another expectation derived from the state of research was that the desired regulatory needs differ between different application areas or scenarios. The fifth hypothesis *The required need for regulation varies between the different use cases* was found to be true, as there was at least one significant difference. Thus, the result is in line with the expectation.

This study shows that there is a difference between the scenarios and further research should investigate which regulatory needs differ in the different scenarios and which regulatory needs are valued the same in each scenario. For a future study, it would therefore be interesting to investigate which regulatory needs are labeled as primary (in each scenario) and which regulatory needs are labeled as secondary (different in scenarios). A possible conjoint study could help identify the relative strengths of individual regulatory needs depending on the different types of benefits AI could provide.

6.2 Answering the Research Question

To answer the research question *RQ: What is the effect of technology acceptance and trust in technology on the need for regulation and how does it impact the perception of fairness of artificial intelligence?*, five different hypotheses were formulated and examined. The first hypothesis was rejected, but findings for the general research question show that the evaluation of the fairness of AI is influenced by the desired regulations. We proposed that the perceptions of fairness are determined by the perceived need for regulation. However, it is equally valid, to assume this association going the other way. With larger samples and using structural equation modeling, the direction of this association could be identified more clearly. We intend to investigate this relationship in more detail in later research.

According to the second hypothesis, there is a negative correlation between technology trust and the need for regulation, where the third hypothesis assumed a positive correlation between trust and perceived fairness. Both were confirmed. The fourth hypothesis was rejected and tells us nothing about the research question. The fifth hypothesis describes that there is an additional difference in regulatory needs between the scenarios of medicine, teaching, and human resource management.

Overall, it follows from the results that technology trust has an influence on the evaluation of the fairness of AI and that technology trust correlates with regulatory needs. In addition, the evaluation of the fairness of AI and the regulatory needs influence each other. According to this study, technology affinity does not correlate with regulatory needs. For further research, it would be interesting to investigate how people generally accept AI and what influence this AI acceptance has on the regulatory needs, and the evaluation of the fairness of AI. In addition, it would be interesting to investigate which regulatory needs are designated as primary and secondary and which other characteristics besides technology trust have an influence on the desired regulatory needs and on the evaluation of the fairness of AI.

7 Conclusion

This paper opened with the observation that fairness in the use of artificial intelligence is one of the central aspects of the public discourse. With the help of a survey, the regulatory needs and expectations for fairness were queried based on concrete scenarios. This study has attempted to answer the question *RQ: What is the effect of technology acceptance and trust in technology on the need for regulation to ensure the fair use of artificial intelligence?*

We confirmed an influence of technology trust on low regulatory needs as hypothesized. Furthermore, it is crucial for the participants to which scenario the need for regulation refers to. It is also interesting to note that the influence of technology trust on the regulatory need for data use was more pronounced than on the regulatory need for data persistence. Based on the results, however, the hypotheses that a high injustice sensitivity leads to a critical evaluation of the fairness of AI and that a low affinity towards technology leads to a high need for regulation had to be rejected.

The reason for the different impact of technology trust compared to technology affinity on regulatory needs should be investigated in further studies. Further results from acceptance research specifically focused on AI are also needed. Different application scenarios are evaluated differently, with high relevance for the practical development of AI. For further research, it is recommended to explore more details about these assessments and to divide the regulatory needs into primary and secondary requirements.

The establishment of new technology depends on the assistance and acceptance of consumers. This study is a first step in understanding the influences on the requirements for fair AI in more detail and can serve as a starting point for further research.

Acknowledgements. This research was supported by the Digital Society research program funded by the Ministry of Culture and Science of the German State of North Rhine-Westphalia. We would further like to thank the authors of the packages we have used.

References

1. Allaire, J., Iannone, R., Presmanes Hill, A., Xie, Y.: distill: 'R Markdown' Format for Scientific and Technical Writing (2020). https://CRAN.R-project.org/package=distill, r package version 1.1
2. Butterwoth, M.: The ICO and artificial intelligence: The role of fairness in the GDPR framework. Comput. Law Secur. Rev.: Int. J. Technol. Law Practice (2018)
3. Calero Valdez, A., Ziefle, M.: The users' perspective on the privacy-utility trade-offs in health recommender systems. Int. J. Hum Comput Stud. **121**, 108–121 (2019)
4. Calero Valdez, A.: hcictools: Tools for data analysis in psychological surveys (2021), r package version 1.1.2
5. Cath, C.: Governing artificial intelligence: ethical, legal and technical opportunities and challenges introduction. Philos. Trans. Roy. Soc. A: Math. Phys. Eng. Sci. **376**(2133) (2018)
6. Cook, D., Augusto, J., Jakkula, V.: Ambient intelligence: technologies, applications, and opportunities. In: Pervasive and Mobile Computing, pp. 277–298 (2007)
7. Fernandez, C., Fernandez, A.: Ethical and legal implications of AI recruiting software. In: ERCIM News, pp. 22–23 (2019)
8. Fuchs, G.: Vertrauen in technische Systeme ist eine Voraussetzung, um Risiken zu bewältigen. https://www.ingenieur.de/karriere/bildung/weiterbildung/vertrauen-ersetzt-wissen/. Accessed 08 July 2019
9. Haladjian, H.H., Montemayor, C.: Artificial consciousness and the consciousness-attention dissociation. Elsevier Inc. (2016)
10. Herzog, L.: Glückwunsch, sie haben die ki überzeugt! https://www.zeit.de/arbeit/2019-05/kuenstliche-intelligenz-arbeitsplatz-fairness-algorithmen-diskriminierung. Accessed 08 July 2019
11. Jantke, K.P.: Informatik und künstliche intelligenz-beiträge zur adaptivität einer kommenden generation intelligenter elearning-systeme. In: eLearning in der Sportwissenschaft: Strategien, Konzeptionen, Perspektiven (eLearning in sports science: Strategies, conceptual design, prospects), pp. 49–70 (2005)
12. Schäffer, U., Weber, J.: Künstliche Intelligenz. Controlling Manag. Rev. **65**(2), 3–3 (2021). https://doi.org/10.1007/s12176-021-0370-0
13. Mainzer, K.: Künstliche Intelligenz - Wann übernehmen die Maschinen?. Springer, Heidelberg (2016)
14. Makridakis, S.: The forthcoming artificial intelligence (ai) revolution: Its impact on society and firms. In: Futures (2017)
15. Miller, D., Brown, E.: Knowing what we know: supporting knowledge creation and sharing in social networks. Am. J. Med. **131**(2) (2018)

16. Rahwan, I.: Society-in-the-Loop: Programming the Algorithmic Social Contract. Springer, Netherlands (2017). https://doi.org/10.1007/s10676-017-9430-8
17. Ray, S., Danks, N.P., Calero Valdez, A.: seminr: Domain-Specific Language for Building and Estimating Structural Equation Models (2021). r package version 2.0.0
18. Renn, O.: Akzeptanzforschung: Technik in der gesellschaftlichen Auseinandersetzung. Chem. unserer Zeit **20**(2), 44–52 (1986)
19. Revelle, W.: psych: Procedures for Psychological, Psychometric, and Personality Research. Northwestern University, Evanston, Illinois (2020). https://CRAN.R-project.org/package=psych, r package version 2.0.12
20. Scherk, J., Pöckhacker-Tröscher, G., Wager, K.: Künstliche Intelligenz - Artificial Intelligence. BMVIT, Bereich Innovation (2017)
21. Schmitt, M., Baumert, A., Fetchenhauer, D., Gollwitzer, M., Rothmund, Schlösser, T.: Sensibilität für ungerechtigkeit. Psychologische Rundschau **60**(1), 8–22 (2009)
22. Schäfer, M., Keppler, D.: Modelle der technikorientierten Akzeptanzforschung - Überblick und Reflexion am Beispiel eines Forschungsprojekts zur Implementierung innovativer technischer Energieeffizienz-Maßnahmen. Zentrum Technik und Gesellschaft (2013)
23. Selker, R., Love, J., Dropmann, D.: jmv: The 'jamovi' Analyses (2020). https://CRAN.R-project.org/package=jmv, r package version 1.2.23
24. Wickham, H., et al.: Welcome to the tidyverse. J. Open Source Softw. **4**(43), 1686 (2019). https://doi.org/10.21105/joss.01686
25. Xie, Y., Dervieux, C., Riederer, E.: R Markdown Cookbook. Chapman and Hall/CRC, Boca Raton, Florida (2020). https://bookdown.org/yihui/rmarkdown-cookbook, ISBN 9780367563837